

GenCore version 5.i.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:24 ; Search time 58.6787 Seconds
(without alignments)
2176.455 Million cell updates/sec

Title: US-09-857-518A-6
Perfect score: 2173
Sequence: 1 MEXIEVSINSKTIKPTSS.....MAMLEQDPHFLALAPKTLI 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1 2373 100.0 452 3 AAB36439 Strawberry
2 2373 100.0 452 3 AAY79656 Strawberry
3 2043 86.1 455 3 AAB36454 Strawberry
4 538.5 22.7 726 3 AAG39696 Arabidops
5 538.5 22.7 823 3 AAG39695 Arabidops
6 538.5 22.7 954 3 AAG39694 Arabidops
7 533.5 22.5 726 3 AAG27611 Arabidops
8 533.5 22.5 823 3 AAG27610 Arabidops
9 533.5 22.5 954 3 AAG27609 Arabidops
10 499 21.0 426 3 AAB36458 Lemon alic
11 498.5 21.0 443 5 ABB92488 Herbicida
12 498 21.0 436 5 ABB91119 Herbicida
13 498 21.0 436 7 ADB91973 Acyltrans
14 495 20.4 446 7 ADB91985 Acyltrans
15 448 18.9 435 7 ADB91986 Acyltrans
16 438 18.5 474 6 ABB82998 P. somnif
17 438 18.5 474 6 ABB82991 P. somnif
18 438 18.5 474 6 ABB82997 P. somnif
19 408.5 17.2 439 2 AAY17403 Clarkia c
20 404.5 17.0 433 2 AAY17400 Clarkia b
21 404.5 17.0 433 6 AAE29783 C. brewer
22 404.5 17.0 433 6 ABB82993 C. brewer
23 404.5 17.0 433 7 ADB91977 Acyltrans
24 397.5 16.8 439 2 AAY17402 Clarkia c
25 387.5 16.3 439 3 AAY51247 C. roseus

RESULT 1

AAB36439

ID AAB36439 standard; protein; 452 AA.

XX

AC AAB36439;

XX

DT 28-FEB-2001 (first entry)

XX

DE Strawberry alcohol acyl transferase protein SEQ ID NO:13.

XX

KW Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase; aromatic ester; alcohol acyl transferase; alcohol dehydrogenase; pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde; alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food; food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt; confectionery; flavouring; oral medication; vitamin; aroma; beverage; alcohol; scent; fragrance; perfume; cosmetic; suspension aid; aluminium salt; anti-perspirant; pharmaceutical; cleaning product; insect pheromone; dye carrier; solvent; insect repellent; miticide; scabicide; plasticiser; deodorant.

XX

OS Fragaria x ananassa.

XX

PN WO200032789-A1.

XX

PD 08-JUN-2000.

XX

PF 02-DEC-1999; 99WO-NL000737.

XX

PR 02-DEC-1999; 98EP-00204018.

XX

PR 12-MAR-1999; 99EP-00200739.

XX

XX (CPRO-) CPRO-DLO CENT PLANTENVERDELINGS REPROD.

PA Aaroni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'connell AP; WPI; 2000-412335/35.

XX

PI N-PSDB; AAC64766.

XX

DR A new DNA sequence encoding a polypeptide with alcohol acyl transferase activity for producing and regulating aromatic and/or aliphatic ester formation in microorganisms, plant cells or plants.

XX

XX Claim 7; Page 71; 163pp; English.

XX

CC The present invention describes nucleotide sequences with thiolase, alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase, aminotransferase and esterase activities, which are involved in the

CC biosynthetic pathway for aliphatic and/or aromatic ester production in
 CC fruit. The nucleotide sequences can be inserted into the genome of a
 CC fruit-producing plant to regulate aliphatic and/or aromatic ester
 CC formation. Aromatic and/or aliphatic esters in microorganisms, plant
 CC cells or plants are produced by inserting thiolase, alcohol acyl
 CC transferase, alcohol dehydrogenase, pyruvate decarboxylase,
 CC aminotransferase and esterase nucleotide sequences into the genome and
 CC feeding the microorganism or plant with alcohol, aldehydes, alpha-keto
 CC acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and
 CC their proteins can be used in the processed food industry as food
 CC additives to enhance the flavour of syrups, ice-creams, frozen desserts,
 CC yoghurts and confectionery. They are used, as flavouring agents for oral
 CC medications and vitamins; provide flavour and aroma in beverages,
 CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or
 CC scent; enhance the flavour or aroma of natural, synthetic or artificial
 CC products; for the production of novel combinations of artificial flavour
 CC substances; as antibacterial or anti-fungal agents; as fragrance or
 CC perfumes in cosmetics, creams, sun-protectant products, hair
 CC conditioners, lengthening agents and fixatives in perfumes, suspension
 CC aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning
 CC products, personal care products and animal care products; as insect
 CC disinfectant additives; as degreasing solvents for electronics; as insect
 CC pheromones; and as dye carriers, solvents, insect repellents, miticides,
 CC scabicides, plasticisers and deodorants. The present sequence represents
 CC the specifically claimed strawberry alcohol acyl transferase
 XX
 SQ Sequence 452 AA;

Query Match 100.0%; Score 2373; DB 3; Length 452;
 Best Local Similarity 100.0%; Pred. No. 3.5e-237;
 Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKIEVSINSKHTIKPSTSTPLQPKYKLTLDQTPPAYVPIVFYPTDHDNLPQTLA 60
 DB 1 MEKIEVSINSKHTIKPSTSTPLQPKYKLTLDQTPPAYVPIVFYPTDHDNLPQTLA 60

QY 61 DLQALSETLTLYYPLSGRVKNLYIDDFEGVPYLEARVNCMTDFLRKKEICLNEFV 120
 DB 61 DLQALSETLTLYYPLSGRVKNLYIDDFEGVPYLEARVNCMTDFLRKKEICLNEFV 120

QY 121 PIKPFMSMEALSDEYPLLGQVNVFDSGIAIGSVSHKLDGGTADCFKLSWGAVFRGR 180
 DB 121 PIKPFMSMEALSDEYPLLGQVNVFDSGIAIGSVSHKLDGGTADCFKLSWGAVFRGR 180

QY 181 ENIIHPSLSAALLFPDRDDLPKXVDQMEALWFGKVKVATRRFVFGKVAISSIQDEAKS 240
 DB 181 ENIIHPSLSAALLFPDRDDLPKXVDQMEALWFGKVKVATRRFVFGKVAISSIQDEAKS 240

QY 241 ESVKPKSRVHVTGFLWKHLIAASRALTS GTTSTLSIAAQAVNLRTRMNMETVLDNATG 300
 DB 241 ESVKPKSRVHVTGFLWKHLIAASRALTS GTTSTLSIAAQAVNLRTRMNMETVLDNATG 300

QY 301 NLFWQAAILLSHTTPEISDLKCDLWNLNGSVKQCGDYFETPKGEGYGMCHYLD 360
 DB 301 NLFWQAAILLSHTTPEISDLKCDLWNLNGSVKQCGDYFETPKGEGYGMCHYLD 360

QY 361 PORTMSSNEPADTYLFSSTWTFNPNLDFGGRTSWIGVAGKISASCKFIILVPTQCGS 420
 DB 361 PORTMSSNEPADTYLFSSTWTFNPNLDFGGRTSWIGVAGKISASCKFIILVPTQCGS 420

QY 421 GIEAWNLEBKAMLEQDPHFLASPKTLI 452
 DB 421 GIEAWNLEBKAMLEQDPHFLASPKTLI 452

RESULT 2
 AAY79656
 ID AAY79656 standard; protein; 452 AA.

XX AAY79656;

AC
 XX
 DT 12-SEP-2003 (revised)
 DT 29-AUG-2003 (first entry)

XX Strawberry alcohol acyl transferase SLE27.
 DE
 XX Strawberry; alcohol acyl transferase; fruit; ripening; ester; flavour;
 KW aroma; transgenic plant.
 XX
 OS Fragaria x ananassa.
 PN EP1006190-A1.
 XX
 PD 07-JUN-2000.
 XX
 PF 02-DEC-1998; 98EP-00234018.
 XX
 PR 02-DEC-1998; 98EP-00204018.
 XX
 PA (CPRO-) CPRO-DLO CNT PLANTENVERDEDELINGS REPROD.
 XX
 PI Verhoeven HA, Van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;
 XX
 DR WPI: 2000-378264/33.
 DR N-PSDB; AAA27666.
 XX
 PT New polynucleotides encoding enzymes from the biosynthetic pathway for
 PT aromatic and/or aliphatic ester production in fruit used to modify plant
 PT flavors.
 XX
 PS Claim 7; Page 60-62; 116pp; English.

XX The present sequence is that of strawberry cv. Elsanta alcohol acyl
 CC transferase SLE27. Expression of SLE27 increases in ripening fruit, with
 CC highest levels in turning and red fruit when volatile ester compounds are
 CC at their maximum. The invention relates to DNA sequences (see AAA27666-
 CC 78) encoding enzymes (see AAY29656-68) involved in the metabolic pathway
 CC leading to the formation of aliphatic and/or aromatic esters in ripening
 CC fruit. The enzymes have alcohol acyl transferase, alcohol dehydrogenase,
 CC pyruvate decarboxylase, thiolase or aminotransferase activity. Expression
 CC vectors comprising the DNA sequences may be used to regulate ester
 CC formation in fruit. Genetically modified plants, plant cells and
 CC microorganisms can be used to produce esters. The DNA sequences,
 CC polypeptides and antibodies are also used to screen fruit: for volatile
 CC ester compounds; for quality such as flavour, fragrance, aroma, scent,
 CC texture or shape; to distinguish between cultivars and varieties; and to
 CC monitor harvest time, post-harvest quality, shelf-life, timing of
 CC pesticide application, and resistance capacity based on volatile ester
 CC profiles. (Updated on 12-SEP-2003 to standardise OS field)

SQ Sequence 452 AA;

Query Match 100.0%; Score 2373; DB 3; Length 452;
 Best Local Similarity 100.0%; Pred. No. 3.5e-237;
 Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKIEVSINSKHTIKPSTSTPLQPKYKLTLDQTPPAYVPIVFYPTDHDNLPQTLA 60
 DB 1 MEKIEVSINSKHTIKPSTSTPLQPKYKLTLDQTPPAYVPIVFYPTDHDNLPQTLA 60

QY 61 DLQALSETLTLYYPLSGRVKNLYIDDFEGVPYLEARVNCMTDFLRKKEICLNEFV 120
 DB 61 DLQALSETLTLYYPLSGRVKNLYIDDFEGVPYLEARVNCMTDFLRKKEICLNEFV 120

QY 121 PIKPFMSMEALSDEYPLLGQVNVFDSGIAIGSVSHKLDGGTADCFKLSWGAVFRGR 180
 DB 121 PIKPFMSMEALSDEYPLLGQVNVFDSGIAIGSVSHKLDGGTADCFKLSWGAVFRGR 180

QY 181 ENIIHPSLSAALLFPDRDDLPKXVDQMEALWFGKVKVATRRFVFGKVAISSIQDEAKS 240
 DB 181 ENIIHPSLSAALLFPDRDDLPKXVDQMEALWFGKVKVATRRFVFGKVAISSIQDEAKS 240

QY 241 ESVKPKSRVHVTGFLWKHLIAASRALTS GTTSTLSIAAQAVNLRTRMNMETVLDNATG 300
 DB 241 ESVKPKSRVHVTGFLWKHLIAASRALTS GTTSTLSIAAQAVNLRTRMNMETVLDNATG 300

QY 301 NLFWAQALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYLD 360
DB 301 NLFWAQALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYLD 360
QY 361 FORTWSSMEPADDIYLFSSWTFNFPDPCGRTSWIGVAGKIESASCKFIILVPTQCGS 420
DB 361 FORTWSSMEPADDIYLFSSWTFNFPDPCGRTSWIGVAGKIESASCKFIILVPTQCGS 420
QY 421 GTEAVNLEEKWAMLEQDPHELALASPKTLI 452
DB 421 GTEAVNLEEKWAMLEQDPHELALASPKTLI 452

RESULT 3
AAB36454
ID AAB36454 standard; protein; 455 AA.
XX
AC AAB36454;
XX

DT 28-FEB-2001 (first entry)
XX
DE Strawberry vesca alcohol acyl transferase protein SEQ ID NO:168.
XX

XX Strawberry: fruit flavour; biosynthetic pathway; aliphatic; thiolase;
KW aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;
KW pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;
KW alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;
KW food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;
KW confectionery; flavouring; oral medication; vitamin; aroma; beverage;
KW alcohol; scent; fragrance; perfume; cosmetic; suspension aid;
KW aluminium salt; anti-perspirant; pharmaceutical; cleaning product;
KW insect pheromone; dye carrier; solvent; insect repellent; miticide;
KW scabicide; plasticiser; deodorant.
XX
OS Fragaria x ananassa.
XX
FN WO200032789-A1.
XX

XX 08-JUN-2000.

XX 02-DEC-1999; 99WO-NL000737.

XX 02-DEC-1998; 98EP-00204018.

XX 12-MAR-1999; 99EP-00200739.

XX (CPRO-) CPRO-DLO CENT PLANTENVERDELINGS REPROD.

XX Aharoni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'connell AP;

XX WPI; 2000-412335/35.

XX N-PSDB; AAC64790.

XX A new DNA sequence encoding a polypeptide with alcohol acyl transferase
PT activity for producing and regulating aromatic and/or aliphatic ester
PT formation in microorganisms, plant cells or plants.
XX
XX Example 5; Page 107; 163pp; English.

XX The present invention describes nucleotide sequences with thiolase,
XX alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,
XX aminotransferase and esterase activities, which are involved in the
XX biosynthetic pathway for aliphatic and/or aromatic ester production in
XX fruit. The nucleotide sequences can be inserted into the genome of a
XX fruit-producing plant to regulate aliphatic and/or aromatic ester
XX formation. Aromatic and/or aliphatic esters in microorganisms, plant
XX cells or plants are produced by inserting thiolase, alcohol acyl
XX transferase, alcohol dehydrogenase, pyruvate decarboxylase,
XX aminotransferase and esterase nucleotide sequences into the genome and
XX feeding the microorganism or plant with alcohol, aldehydes, alpha-keto
XX acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and
XX their proteins can be used in the processed food industry as food
XX additives to enhance the flavour of syrups, ice-creams, frozen desserts,
XX yoghurts and confectionery. They are used: as flavouring agents for oral

CC medications and vitamins; provide flavour and aroma in beverages,
CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or
CC scent; enhance the flavour or aroma of natural, synthetic or artificial
CC products; for the production of novel combinations of artificial flavour
CC substances; as antibacterial or anti-fungal agents; as fragrance or
CC perfumes in cosmetics, creams, sun-protectant products, hair
CC conditioners, lengthening agents and fixatives in perfumes, cleaning
CC aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning
CC products, personal care products and animal care products; as
CC disinfectant additives; as degreasing solvents for electronics; as insect
CC pheromones; and as dye carriers, solvents, insect repellents, miticides,
CC scabicides, plasticisers and deodorants. The present sequence represents
CC strawberry vesca alcohol acyl transferase, from the present invention
XX
XX SQ Sequence 455 AA;

Query Match 86.1%; Score 2043; DB 3; Length 455;
Best Local Similarity 87.6%; Pred. No. 7.1e-203;
Matches 397; Conservative 15; Mismatches 39; Indels 2; Gaps 2;

QY 1 MEKIEVINSKTIKIPSTSTPQPKYKLLDQLTTPAYVPIVFFYPTDHD-FNLPTQL 59
DB 1 MEKIEVSIISKTIKIPSTSSPLQPKYKLLDQLTTPSYVVMVFFYPTGPAVFNL-QTL 59
QY 60 ADLRQALSETLTLXVPLSGRVKNLYIDDFEGVPYLEARVNCMDTDFLRKIECINEF 119
DB 60 ADLRHALSETLTLXVPLSGRVKNLYIDDFEGVPYLEARVNCMDNDFELPKIECINEF 119
QY 120 VPIKPSMEAISDERYPPLLGQVNWETSGIAGVSVSHKLLIDGGTADCFLKSGCAVPRGC 179
DB 120 VPIKPSMEAISDERYPPLLGQVNIENSGIAGVSVSHKLLIDGRTSDCFLKSCAVFRGS 179
QY 180 RENIHPISLEAALLPDPDDIPEKYVDOMEALWFAKKVATRRFVGKAISSIODEAK 239
DB 180 RDKIHPNLSQALLPDPDDIPEKYARQMEGLWFGKKVATRRFVGKAISSIODEAK 239
QY 240 SESVPKPSRVHVTGFLWKHLIAASRALTSGITSTRLSIAAQAVNLRNMTVTDNAT 299
DB 240 SESVPKPSRVQVTSFLWKHLIATSRLTSGITSTRLSIATQVNVISRNMTVMDNAI 299
QY 300 GNLFWQAQALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETFKGEGYGRMCEYL 359
DB 300 GNLFWPAPALLESHTTPEISDLKCDLVNLLNGSVKQCNQDYPFETPMGEGYGSMEYL 359
QY 360 DPORTWSSMEPADDIYLFSSWTFNFPDPCGRTSWIGVAGKIESASCKFIILVPTQCG 419
DB 360 DPORTWSSMEPADDIYLFSSWTFNFPDPCGRTSWIGVAGKIESAFCLNLTILVPTPCD 419
QY 420 SGTEAVNLEEKWAMLEQDPHELALASPKTLI 452
DB 420 TGTEAVNLEEKWAMLEQDPHELALASPKTLI 452

RESULT 4

AAG39696

ID AAG39696 standard; protein; 726 AA.

XX AAG39696;

AC AAG39696;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 49156.

DE Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439. 99US-01218252P. 99US-01429202P.
XX 25-FEB-1999; 99US-01218252P. 99US-01429277P.
PR 03-MAR-1999; 99US-01233180P. 99US-0143542P.
PR 09-MAR-1999; 99US-01233180P. 99US-0143624P.
PR 23-MAR-1999; 99US-0125788P. 99US-0144005P.
PR 23-MAR-1999; 99US-0125788P. 99US-0144085P.
PR 29-MAR-1999; 99US-0126264P. 99US-0144086P.
PR 01-APR-1999; 99US-0126785P. 99US-0144325P.
PR 01-APR-1999; 99US-0127452P. 99US-0144331P.
PR 06-APR-1999; 99US-0128234P. 99US-0144332P.
PR 08-APR-1999; 99US-0128711P. 99US-0144333P.
PR 16-APR-1999; 99US-0129845P. 99US-0144334P.
PR 19-APR-1999; 99US-0130077P. 99US-0144335P.
PR 21-APR-1999; 99US-0130449P. 99US-0144352P.
PR 23-APR-1999; 99US-0130510P. 99US-0144632P.
PR 23-APR-1999; 99US-0130891P. 99US-0144884P.
PR 28-APR-1999; 99US-0131449P. 99US-0144814P.
PR 30-APR-1999; 99US-0132048P. 99US-0145086P.
PR 30-APR-1999; 99US-0132407P. 99US-0145088P.
PR 04-MAY-1999; 99US-0132484P. 99US-0145085P.
PR 05-MAY-1999; 99US-0132485P. 99US-0145087P.
PR 06-MAY-1999; 99US-0132486P. 99US-0145089P.
PR 06-MAY-1999; 99US-0132487P. 99US-0145132P.
PR 07-MAY-1999; 99US-0132863P. 99US-0145145P.
PR 11-MAY-1999; 99US-0134256P. 99US-0145218P.
PR 14-MAY-1999; 99US-0134218P. 99US-0145224P.
PR 14-MAY-1999; 99US-0134219P. 99US-0145276P.
PR 14-MAY-1999; 99US-0134221P. 99US-0145913P.
PR 14-MAY-1999; 99US-0134370P. 99US-0145918P.
PR 18-MAY-1999; 99US-0134768P. 99US-0145919P.
PR 19-MAY-1999; 99US-0134941P. 99US-0145951P.
PR 20-MAY-1999; 99US-0135124P. 99US-0146386P.
PR 21-MAY-1999; 99US-0135353P. 99US-0146388P.
PR 24-MAY-1999; 99US-0135629P. 99US-0146389P.
PR 25-MAY-1999; 99US-0136021P. 99US-0147038P.
PR 27-MAY-1999; 99US-0136392P. 99US-0147204P.
PR 28-MAY-1999; 99US-0136782P. 99US-0147302P.
PR 01-JUN-1999; 99US-0137222P. 99US-0147192P.
PR 03-JUN-1999; 99US-0137528P. 99US-0147260P.
PR 04-JUN-1999; 99US-0137502P. 99US-0147303P.
PR 07-JUN-1999; 99US-0137724P. 99US-0147416P.
PR 08-JUN-1999; 99US-0138094P. 99US-0147433P.
PR 10-JUN-1999; 99US-0138540P. 99US-0147935P.
PR 10-JUN-1999; 99US-0138847P. 99US-0148171P.
PR 14-JUN-1999; 99US-0139119P. 99US-0148319P.
PR 16-JUN-1999; 99US-0139452P. 99US-0148341P.
PR 16-JUN-1999; 99US-0139453P. 99US-0148565P.
PR 17-JUN-1999; 99US-0139482P. 99US-0148684P.
PR 18-JUN-1999; 99US-0139484P. 99US-0149368P.
PR 18-JUN-1999; 99US-0139455P. 99US-0149175P.
PR 18-JUN-1999; 99US-0139456P. 99US-0149426P.
PR 18-JUN-1999; 99US-0139457P. 99US-0149722P.
PR 18-JUN-1999; 99US-0139458P. 99US-0149723P.
PR 18-JUN-1999; 99US-0139459P. 99US-0149929P.
PR 18-JUN-1999; 99US-0139460P. 99US-0149902P.
PR 18-JUN-1999; 99US-0139461P. 99US-0149930P.
PR 18-JUN-1999; 99US-0139462P. 99US-0150566P.
PR 18-JUN-1999; 99US-0139463P. 99US-0150884P.
PR 18-JUN-1999; 99US-0139750P. 99US-0151065P.
PR 18-JUN-1999; 99US-0139763P. 99US-0151066P.
PR 21-JUN-1999; 99US-0139817P. 99US-0151080P.
PR 22-JUN-1999; 99US-0139899P. 99US-0151303P.
PR 23-JUN-1999; 99US-0140353P. 99US-0151438P.
PR 23-JUN-1999; 99US-0140354P. 99US-0151930P.
PR 24-JUN-1999; 99US-0140695P. 99US-0152363P.
PR 28-JUN-1999; 99US-0140823P. 99US-0153070P.
PR 29-JUN-1999; 99US-0140991P. 99US-0153756P.
PR 30-JUN-1999; 99US-0141287P. 99US-0154018P.
PR 01-JUL-1999; 99US-0141842P. 99US-0154039P.
PR 01-JUL-1999; 99US-0142154P. 99US-0154779P.
PR 02-JUL-1999; 99US-0142055P. 99US-0155139P.
PR 06-JUL-1999; 99US-0142390P. 99US-0155486P.
PR 08-JUL-1999; 99US-0142803P. 99US-0155559P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142927P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147433P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153756P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.


```
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157111P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160788P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161982P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.78; Score 538.5; DB 3; Length 726;
Best Local Similarity 33.67; Pred. No. 3.7e-46;
Matches 153; Conservative 88; Mismatches 178; Indels 37; Gaps 15;

QY 3 KIEVINSKHTIKPSTSTP--LQPYKLTLLDQTPPAYVPIVFYPTIDHDFNLPTOLA 60
Db 284 QMEVDVVSRIILKPS-SPTNHLKKFKLSLEQLGPIFGPWVFYS-ANNSIKPTQLQ 341
QY 61 DLQALSETLTYPLSGRVNNLYIDDFFGVYPLEARNVCMDFLRKIKIECLNEFV 120
Db 342 MLKKSLSLTHFYPLAGRLKGNISICNDSGADFLARNVSPLSNLLLEPSSDSLQOLI 401
QY 121 PIKPFSEALSDERYPLGVQNVFDSG-TAIGVSVSHKLIDGTADCFKLSGAV-FRG 178
Db 402 PTVSDSTET---RTRLLAQASFEGCSMSIGVICSHKLADATSIGLFMKSWAAISRG 457
QY 179 CRENIHPSLSEALPPP---RDLPEKYVDQMEALWPAKKVATRRFVFGKAISSIQ 235
Db 458 SIKTIGAP-VFDVTKIPPGNPFSTSPAPVVEPIWM---NQTLKRFIFDSSSIQALQ 512
QY 236 DEAKSEVPKPSRVHATGFLWKLHIAASRALTSGTSTRLSIAAQNLRTRNMETVL 295
Db 513 AKASSEVNPQTRVRAVSALIWKSAMKATRTV-SGT--SKPSILANSVLSRSVSPPT 568
QY 296 DNATGKLFWAQLLESHITPETSDIKLDLVNLLGSKQNGDYFFETPKGEGYGR- 354
Db 569 KNSIGNLVSTPAKAE-----EGINQTKLQTLVSKIRKAKQRFDIHIPKLVGNPNATEI 623
QY 355 MCBYLPQRTMSSMEPAP--DIYLFSSWTFN-FNELDFMGRTSWIGVAGKIESACKFI 411
Db 624 ICSTY---QKEAGDMIASGDFDFFISACREGLYETDFGKPVWVGFPSVRQK---NIV 677
QY 412 ILVPTQCGSGIEAWNVLEBKAMLEQDPHPLAAS 447
Db 678 TLLDTKAGGIEAWNVLEBKAMLEQDPHPLAAS 713
```

RESULT 5
AAG39695
ID AAG39695 standard; protein; 823 AA.
XX
AC AAG39695;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49155.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135628P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.

```
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.

Query Match 22.7%; Score 538.5; DB 3; Length 823;
Best Local Similarity 33.6%; Pred. No. 4.5e-46;
Matches 153; Conservative 88; Mismatches 178; Indels 37; Gaps 18;

QY 3 KIEVSINSKHTIKPSTSTP--LQPKLGLDQUTPPAYVPIVFPYITDHFNLPTQLA 60
Db 381 QMRVDVVSRIIKPS-SPTPNELKKFKLSLLEQLGPTIFGWMVFYS-ANNSIKPTEQLQ 438
QY 61 DLROALSETLTYPLDLSGRVKNLYIDDFEGVYLKARVNCMTDFLRLKIECLNEFV 120
Db 439 MLKSLSETLTHFYPLAGLAKENISIDCNDSGADFLKARVNSPLNLLLEPSDSLQOLI 498
QY 121 PIKPFMEAISCDERYPLLGQVKNVDSG-IAIGVSVSHKLDGGTADCFKLSWGA-PRG 178
Db 499 PTVSDSIET-----RTRLLAQSPFCGSMISGVCISHKLDATSIGLFMKSNIAISRG 554
```


PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150894P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.7%; Score 538.5; DB 3; Length 954;
Best Local Similarity 33.6%; Pred. No. 5.7e-46;
Matches 153; Conservative 88; Mismatches 178; Indels 37; Gaps 18;

Qy 3 KIEVSINSKHTIKPSTSTP--LQPYKLTLLDTPPAYVPIVFEYPIITODHFNLPQTLA 60
Db 512 QMRVDVVSRDIIKPS-SPTNHLKKFKLSLLEQGTIFGPMVFYIS-ANKSIKPTCOLQ 569
Qy 61 DLROALSETLTVYPLSGRVKNMLYIDDFEGVYLEARNVCOMTDIFLEFKICLNEFV 120
Db 570 MLKKSLSLTHFYPLAGRLKGNISIDCNSGDGFLEARNVSNLLEPSSLSLOQLI 629
Qy 121 PIKPFMEALSDEYFILLGVQNVFDSG-IAIGVSVSHKLIDGTADCFLKSGAV-FRG 178
Db 630 PTVSDSIET---RTRLLAQASFEGCSMSIGVICISHKLADA--SIGLEMKSWAIIISRG 685
Qy 179 CRENIHPSLSSEALLPPP---RDDLPKVDQMEALWFAGKVATRRFVFGKAISSIQ 235
Db 686 SIKTIGAP-VFDTVKIPFGNPFSETSPVVEPIMK---NOTLSRRFIPDSSSIQALQ 740
Qy 236 DEAKSEVPKPSRVHAVTGFLLWKHLIAASRALTSQTTSTRLSTIAAQAVNLRNMETVL 295
Db 741 AKASSFEVQPTRAVEAVSALIKSAMKATRV-SGT--SKESILANSVLSRSVY-PPFT 796
Qy 296 DNATGNLFWAQAILESHHTPEISDLKDLVNLKNGVKQNGDYFETPKGEGYGR- 354
Db 797 KNSIGNLVSYFAAKAE-----EGINOTKQLTLVSKIRKAKORFERDIHIPKLVGNPNATEI 851
Qy 355 MCEYLDFORTMSSMEPAP--DIYLFSSWTF-FNPLDFGMRGRTSWIGVAGKIESASCFI 411
Db 852 ICSTY---QKAGDMIASGDFDFYIFSSACRFLGYETDFGKGPVWVGPPSVRQK---NIV 905
Qy 412 ILVPTQCGSGIEAWVNLEEKAMLEQDPHPLALAS 447
Db 906 TLLDTKEAGGIEAWVNLEQDMNLFQDRELLQFAS 941

RESULT 7
AAG27611
ID AAG27611 standard; protein; 726 AA.
XX AC AAG27611;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 32516.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KM hybridisation assay; Genetic mapping; gene expression control; promoter;
XX KM termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 03-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 25-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127452P.
XX PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130044P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 26-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 04-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 16-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 09-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149932P.
PR 23-AUG-1999; 99US-0149932P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.

```
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.5%; Score 533.5; DB 3; Length 726;
Best Local Similarity 33.3%; Pred. No. 1.2e-45;
Matches 152; Conservative 87; Mismatches 180; Indels 37; Gaps 18;

Qy 3 KIEYSINKHTIKESTSTP--LOPYKLTLDLTTPPAYVIVPYPITDHDNLPQTLA 60
Db 284 QMRVDVSRDRIKES-SPTNHLKKFKLSLEQLGPTIFGPMFPFYS-ANNSIKPTQLQ 341
Qy 61 DLQALSETLTYPLSGVRKNLYIDFERGVPIELARVNCMDTDFURLAKIECLNEFV 120
Db 342 MLKKSLETLTHFYPLAGLKENISIDCNDSGADFLERVNSPLSSLLPESSLSLOQLI 401
Qy 121 PIKPSMEASIDERNPLIGVQNVFDSG-IAIGSVSHKLIDGGTADCFKLSWGAV-FRG 178
Db 402 PTVSDSIET-----RTRLLAQASFECGSMGIVCISHKLADATSIGLFMSNAIISRG 457
Qy 179 CRENIHPSLSEALLPDP---RDDLPKYVDQMEALFAGKVKVATRFVGVKAISSIQ 235
Db 458 SIKTIGAP-VDTWKIPFGNPFSETSPAPVPEPEIMM---NQTLSKRFIFDSSIQALQ 512
Qy 236 DEAKSEVPKPSRVHVTGFLWKHLIAASRALTSQTSTRLSIAQAQVNLRTMNMETVL 295
Db 513 AKASSPEVNPQTRVEAVSALIKSMKATRV-SGT--SKPSILANSASLSRVSP-PPT 569
Qy 296 DNATGNLFWQAQIIELESHHTPTREISDLKDLVNLINGSVKQNGDYFETPKGEGYGR- 354
Db 569 KNSIGNLVSPFAKAE-----EGINQTKLQTLVKIRKAKQRFROIHIPKLVGNPNATEI 623
Qy 355 MCEYLDQRTWSSMEAP--DIYLFSSWTNF-FNPLDFGWTGRTSWIGAGKIESASCKFI 411
Db 624 ICSTY---QKAGDMIAGDFFIYFSSACRGLVETDFGKGVKGVWVGPPSVRQK---NIV 677
Qy 412 ILVPTQCGSGIEAVNLEERKQWALEQDPHFALAS 447
Db 678 TLLDTKEAGGIEAVNLEERKQWALEQDPHFALAS 713

RESULT 8
AAG27610
ID AAG27610 standard; protein; 823 AA.
XX AC AAG27610;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 32515.
```

```
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125783P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139877P.
PR 22-JUN-1999; 99US-0139899P.
```

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

99US-0121825P.

99US-0123180P.

99US-0123548P.

99US-0125783P.

99US-0126264P.

99US-0126785P.

99US-0127462P.

99US-0128234P.

99US-0128714P.

99US-0129845P.

99US-0130077P.

99US-0130449P.

99US-0130510P.

99US-0130891P.

99US-0131449P.

99US-0132048P.

99US-0132407P.

99US-0132484P.

99US-0132485P.

99US-0132486P.

99US-0132487P.

99US-0132863P.

99US-0134256P.

99US-0134218P.

99US-0134219P.

99US-0134221P.

99US-0134370P.

99US-0134768P.

99US-0134941P.

99US-0135124P.

99US-0135353P.

99US-0135629P.

99US-0136021P.

99US-0136392P.

99US-0136782P.

99US-0137222P.

99US-0137528P.

99US-0137502P.

99US-0137724P.

99US-0138094P.

99US-0138540P.

99US-0138847P.

99US-0139119P.

99US-0139452P.

99US-0139453P.

99US-0139492P.

99US-0139454P.

99US-0139455P.

99US-0139456P.

99US-0139457P.

99US-0139458P.

99US-0139459P.

99US-0139460P.

99US-0139461P.

99US-0139462P.

99US-0139463P.

99US-0139750P.

99US-0139763P.

99US-0139877P.

99US-0139899P.

```
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140696P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142300P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 18-JUL-1999; 99US-0144005P.
PR 18-JUL-1999; 99US-0144280P.
PR 18-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145818P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.

PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 22.5%; Score 533.5; DB 3; Length 823;
Best Local Similarity 33.3%; Pred No. 1.5e-45;
Matches 152; Conservative 87; Mismatches 180; Indels 37; Gaps 18;

QY 3 KIEVINSKHTIKPSTSTP--LQPYKLTLLDQITPPAYVPIVFFYFPIITDHDNLPOTLA 60
Db 381 QMRVDVVSRIKPS-SPTNHLKKFKLSLEQLGPTIFGPMVFFYS-ANNSIKPTQLQ 438
QY 61 DLROALSETLTYPLSGRVNNLYIDDPGPGVPLEARVNCMDTPIRLKKECLNEFV 120
Db 439 MLKKSLSLTHEYPLAGRLKGNISIDCNDSGADFLFARVNSPLSLLEPSSDSLOQLI 498
QY 121 PIKPFMEALSDEERYPLGVQVNVDSG-TAIGVSVSHKLIDGTDACELKSMGAV-FRG 178
Db 499 PTVDSIET----RTRLLAQASFFCGSMSIGVSHKLADATSIGLFKSWAAISR 554
QY 179 CRENIHPSLSEALLPPP---RDDLPEKYVDQMEALWPAKKVATRFVFGVKAISIQ 235
Db 555 SIKTIGAP-VFDTVKIFPPGPNFSETSPAPVVEPIYX---NQTLSKRFIDSSSIQALQ 609
QY 236 DEAKSEVPKPSRVHVTGFLWKHLIAASRALTSGTSTRLSTAAQAVNLRNMNMTVL 295
Db 610 AKASSPEVNQPTREAVSALIKWSAMKATRV-SGT--SKPSTIANASLSRVS-PPFT 665
QY 296 DNATGNLFWQAQILSLSHTTPTISDLKCDLVNLLNGSVKQCNGDYFETFKGEGYGR- 354
```

```
Db 666 KNSIGNLVSYFAKAE-----EGINTQKLTQVSKIRKAKQRFROIHIPKLVGNPNATRI 720
QY 355 MCEYLDQFQRTMSMEBPAP--DIYLFSSWTFN-FNPLDFQWGRTSWIGVAGKTESASCKXI 411
Db 721 ICSY---QKEAGDMIASGDFDFYITSSACRFGLYETDFQWGRPVWVGFFSVQK---NIV 774
QY 412 ILVPTQCGSGIEAWNLSEKAMLEQDPHFALAS 447
Db 775 TLLDTKAGIEAWNLSEKAMLEQDPHFALAS 810

RESULT 9
AAG27609
ID AAG27609 standard; protein; 954 AA.
XX
AC AAG27609;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32514.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126284P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140351P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144864P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145152P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
```


Db 287 LLAESLIGNIM-FSSVVLGIS-----QOEIKIEKAVRDLR---KQ--GDDLQNVIKBEQG 336
 Qy 352 -----YGRCEYLDLDFORTWSSMEPADDIYLFSSWTFFNPL----DFGW 391
 Db 337 SSSAMIGSKLANLMTNYSKL-----SYETHEP-----YTVSSWCKL--PLYEASFGW 382
 Qy 392 GRTSWIGVAGKIESASCKFIILVPTQCGSGIEAWNVLEEKAMLEQDPHFLALAS--PK 449
 Db 383 GSPWVI--AGNVAPMLENVMLDSKDGQIEAFVTLPEENMLSLQNTTELLAFASVNP 440
 Qy 450 TLI 452
 Db 441 VLV 443

RESULT 12
 ABB91119
 ID ABB91119 standard; protein; 436 AA.
 XX AC ABB91119;
 XX 31-MAY-2002 (first entry)
 XX Herbicidally active polypeptide SEQ ID NO 330.
 XX Herbicidal; plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX WQ200210210-A2.
 XX 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX (PARB ↓ BAYER AG.
 XX Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.
 XX Claim 5; SEQ ID NO 330; 261pp + Sequence Listing; English.
 XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides
 XX Sequence 436 AA;
 SQ

Query Match 21.0%; Score 498; DB 5; Length 436;
 Best Local Similarity 30.9%; Pred. No. 2.7e-42;
 Matches 145; Conservative 90; Mismatches 184; Indels 50; Gaps 19;

Qy 1 MEKIEVSINSKHTIKESTSTP--LQPYKLTLLDQLTPPAYVPIVFFYPITDHD---FNL 55
 Db 1 MEK-NVEILLSREIVKPS-SPTDDKRLNLSLLDLSSEPMYTGALLFYADPQNLGFS 58
 Qy 56 PQLADIQALSTLTLYPLSRGVKNLVLDIFEGVPLVLEARNVCDMTDFLRKIEC 115
 Db 59 EETSLKMKSLSKLTPIFYPLAGRIIGS-FVECNDEGAVFIARVDHLLSEFLKCPVPS 117

Qy 116 LNEFVPIKPSMEATISDERYPILGWQWVFD-SGIAIGVSVSHKLIDGTADCFKSWGA 174
 Db 118 LELLIPVAKREAVT---WPVLLIQANFFSCGGLVITICVSHKITDATSLAMFIRGMAE 174
 Qy 175 VFRGRENIIIPSLSEAAALLPPRDCLPEKXVDOMEALWFAGKKVATREFFVGVKAISII 234
 Db 175 SSRGIGITLI-PSFTASEVFPKPLDELSPKMDRAEEV--EEMSCVTKEFVFDASKIKL 231
 Qy 235 QDEAKESVPKPSRVHVTGFLWKHLIAASRALTSGLTSTRLSIAAOAVNLTRMMETV 294
 Db 232 RAKASRNLVKPNTRVEAVTALFWRCVTKVSL---SSLTPRTSVLQILVNLGRK--VDSL 286
 Qy 295 LDNATGNLFWN-----AQALELSHTTPEISDLKCLDLVNLNGSVKQCNDYFETPK- 347
 Db 287 CENTIGNMLSLMLKNEEAAIERIQDVVDEIRRAK--BIFSLNCKEMSKSSRIFFELLE 344
 Qy 348 -GREGYGRMCEYLDLDFQRTWSSMEPADDIYLFSSWTNF--FNPLDFQWGRTSWTGVAGKIES 405
 Db 345 IGKV-YRGENEM-----DLWMSNCKLGLYDADFQWGPVW--VTGRGTS 387
 Qy 406 ASCKPIILVPTQCGSGIEAWNVLEEKAMLEQDPHFLALAS--PKTLI 452
 Db 388 HFKNMLMLIDTKDGEIEAWITLTERQMSLFECQDELLESASLNPVLI 436

RESULT 13
 ABB91973
 ID ABB91973 standard; protein; 436 AA.
 XX AC ABB91973;
 XX 04-DEC-2003 (first entry)
 XX Acyltransacylase enzyme #11.
 XX Acyltransacylase; transacylase; paclitaxel; taxoid;
 XX paclitaxel biosynthetic pathway; enzyme.
 XX Arabidopsis thaliana.
 XX US2003108891-A1.
 XX 12-JUN-2003.
 XX 18-SEP-2002; 2002US-00166984.
 XX 30-SEP-1999; 99US-00411145.
 XX 07-DEC-1999; 99US-00457046.
 XX 25-MAY-2001; 2001US-00866570.
 XX (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX Croteau RB, Walker KD, Schoendorf A, Wildung WR;
 XX WPI; 2003-659103/62.
 XX New purified protein useful for producing paclitaxel and related taxoids.
 XX Example 1; Fig 6; 135pp; English.
 XX The invention relates to a transacylase enzyme and the polynucleotide
 CC encoding it. The invention also relates to a recombinant nucleic acid
 CC molecule comprising a promoter sequence operably linked to the
 CC polynucleotide encoding the transacylase enzyme and a cell transformed
 CC with the recombinant nucleic acid. The transacylase enzyme is useful for
 CC producing paclitaxel, related taxoids and intermediates in the paclitaxel
 CC biosynthetic pathway. This sequence represents an acyltransacylase enzyme
 CC of the invention.
 XX Sequence 436 AA;
 SQ

Query Match 21.0%; Score 498; DB 7; Length 436;
 Query Match 21.0%; Score 498; DB 7; Length 436;

```

Best Local Similarity 30.9%; Pred. No. 2.7e-42;
Matches 145; Conservative 90; Mismatches 184; Indels 50; Gaps 19;

QY 1 MEKIEVSNKHTTKPSSSTP--LQPKYKLLDQLTTPAYVPTVFFVPIETHD---FNL 55
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1 MEK-NVEILSREIVKPS-SPTPDKRIILNLSILDLSSPYTGALLFYAADPQNLGFST 58
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 56 PQTLDADLQALSETILYPLSGRVKNLYIDDFEEGVPLYEARVNCMTDFLRLKIC 115
  :|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 59 BETSLKLSKSLKTIPIFYPLAGRIIGS-FVECNDEGAVFTEARVDHLLSEFLKCPVPS 117
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 116 LNEFPIKPFMEALISDRYPLLVGVNVFD-SGIAIGSVSHKLIDGGTADCFLKSWG 174
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 118 LELLIPVEAKSREAVT---WPLLIIQANFPSCGSLVITICVSHKITDATSLAMFIRGAE 174
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 175 VFRGCRNIIHPSISEAALLPPPRDDIPEKYVDQWEALWFAKKVATRRFVGVKAISSI 234
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 175 SSRGLIGTLI-PSPTASEVFPKPIDELSPKPMDSKEEV--EEMSCVTKRFVFDASKIKL 231
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 235 QDEAKSSVPKPSVHVATGFLWKHLIRASRLTSGTSTFSLTAAQAVNLRTMNMETV 294
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 232 RAKASRLNVKNPTRVEAVTALFWCVTKVSEL-----SSLTPRTSVIQLIVNLRGK--VDSSL 286

QY 295 LDNATGNLFW-----AQAILLSHTTPEISDLKCLDLVNLINGSVKQCNQDYFETFK- 347
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 287 CENTIGNMLSLMLKNEBAALIERIQDVVDERRAK--EIFSLNCKEMSKSSRIFELLE 344

QY 348 -GKEGYGRMCEYLLDFQRTMSMEPADIYLFSSWTFN-FNPLDPGWRGRTSWIGVAKKIES 405
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 345 IGKV-YORGEM-----DLWMSNSWCKLGLYDADFGMGKPYW--VTGRGTS 387

QY 406 ASCKFIILVPTQCGSGIEAWNVLEEFKQWMLQDPHPFLALAS--PKTLI 452
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 388 HFKNMLLIDTKDGEIEKAWITLTTEQMSLPECDQELLESASLNPPVLI 436

RESULT 14
ADB91985
ID ADB91985 standard; protein; 446 AA.
XX
AC ADB91985;
XX
DT 04-DEC-2003 (first entry)
XX
DE Acyltransacylase enzyme #23.
XX
KW Acyltransacylase; transacylase; pacilitaxel; taxoid;
KW pacilitaxel biosynthetic pathway; enzyme.
XX
OS Arabidopsis thaliana.
XX
US2003108891-A1.
XX
PD 12-JUN-2003.
XX
PF 18-SEP-2002; 2002US-00165984.
XX
PR 30-SEP-1999; 990S-00411145.
PR 07-DEC-1999; 99US-00457046.
PR 25-MAY-2001; 2001US-00866570.
XX
(PUNW ) UNIV WASHINGTON STATE RES FOUND.
PA
XX
PI Croteau RB, Walker KD, Schoendorf A, Wildung MR;
XX
WPI; 2003-659103/62.
XX

```

molecule comprising a promoter sequence operably linked to the polynucleotide encoding the transacylase enzyme and a cell transformed with the recombinant nucleic acid. The transacylase enzyme is useful for producing paclitaxel, related taxoids and intermediates in the paclitaxel biosynthetic pathway. This sequence represents an acyltransacylase enzyme of the invention.

XX

SQ Sequence 446 AA;

Query Match
Best Local Similarity 20.4%; Score 485; DB 7; Length 446;
Matches 143; Conservative 91; Mismatches 165; Indels 92; Gaps 23

QY 4 TEVSIKSHITKPSTSTP--LQPKYLRLDQLTPPAYVIPVFYPITDHD-FMLPQTLLA 60
DB ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 6 MKVETISKEIIKPS-SPTPNNLQTLQLSDIYHILPPVTVAFLFY--TKNDLSIQSEHTS 62

QY 61 DIRQALSELTLIXPLSGRKNLVIDDFERGVPLEARV-NCDMTDFELRLKIERCLAEF 119
DB ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 63 KKLTSLETLNKFYDLAGRI-TGTVDCIDEGAFVDARVNCPTEFLKPCPDFDALQOL 121

QY 120 VPIKPFMEALSDEIRYPVLGGVVNVFD-SGIAIGVSVSKHLIDGGTADCFLKSWGAVFRG 178
DB :||:|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 122 LPLDVVDNPYYAAATWPLLWKATYFGCGGAIGICITHKIADAASISTFTIRSWAATAARG 181

QY 179 CREN---LIHSLSSEAALLFPRDDLPKYVDQEMALWFAGKVVA-TREFFVGUKAISGI 234
DB |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 182 --ENDAAAMESPVFAGANFYIPANEAKLPDAEQ-----AKRSITIKFEVFEASKVEDL 234

QY 235 QDEAKS-EIVPKPRGRHVAVTGFLWKHLIAASRALTSGETTSLSTIAAQAVNLRTMKMYET 293
DB :||:|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 235 RTKAASEBTVDPQRVESVTALIWKCFVASSK----ITTCDHKVLVQLANLRSKI-PSL 288

QY 294 VLDRATGNLFWWA-----QAILELSTHTPEISDLKLC D----- 326
DB :||:|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 289 LOESSIGNIMFSFVSVLIGRGGEVKIVEAVNRDLRAKKBELGTVIDEGGSSDSMWISK 348

QY 327 LVNLNGSVQCNGDYPTFKKGEGYGRCMCEYLDFOQTMSSWEPA PD IYLFSSWTNFNP 386
DB ||||:-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-|-
 349 LANLM-----LTNYSEL-----SYETHPE-----YTYSWSCKL--P 377

QY 387 L---DFWGRTSWTGVAGKTIESACKFIILVPOTCGSIGIEAWNLEEEKMAWLQDPHF 443
DB |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
 378 LYEAASFWDSPFW-VVGNVSPVLGNLANLIDSXOGGIEFAFVTLPEENMSSFQNPELL 435

QY 444 ALA--SPKTLII 452
DB |||:~::~|||:
 436 AFATWNPSVLV 446

RESULT 15
ADB91986
ID ADB91986 standard; protein; 435 AA.
XX
AC ADB91986;
XX
DT 04-DEC-2003 (first entry)
XX
DE Acyltransacylase enzyme #24.
KW Acyltransacylase; transacylase; paclitaxel; taxoid;
KW paclitaxel biosynthetic pathway; enzyme.
OS Arabidopsis thaliana.
XX
FN US2003104891-A1.
XX
PD 12-JUN-2003.
XX
PF 18-SEP-2002; 2002US-00166984.
XX
FR 30-SEP-1999; 99US-00411145.
FR 07-DEC-1999; 99US-00457046.

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	498	21.0	436	3	US-09-457-046B-72		Sequence 62, Appl
2	485	20.4	446	3	US-09-457-046B-74		Sequence 74, Appl
3	448	18.9	435	3	US-09-457-046B-60		Sequence 60, Appl
4	404.5	17.0	433	3	US-09-457-046B-66		Sequence 66, Appl
5	387.5	16.3	439	3	US-09-457-046B-68		Sequence 68, Appl
6	308	13.0	331	3	US-09-457-046B-59		Sequence 59, Appl
7	242	10.2	461	3	US-09-457-046B-64		Sequence 64, Appl
8	230	9.7	448	1	US-08-207-904-2		Sequence 2, Appl
9	230	9.7	448	1	US-08-207-904-17		Sequence 17, Appl
10	228.5	9.6	451	3	US-09-457-046B-69		Sequence 69, Appl
11	225	9.5	445	3	US-09-457-046B-73		Sequence 73, Appl
12	222	9.4	497	3	US-09-457-046B-65		Sequence 65, Appl
13	219	9.2	482	3	US-09-457-046B-63		Sequence 63, Appl
14	179	7.5	443	3	US-09-457-046B-50		Sequence 50, Appl
15	178.5	7.5	461	3	US-09-457-046B-70		Sequence 70, Appl
16	177.5	7.5	441	3	US-09-457-046B-54		Sequence 54, Appl
17	177.5	7.5	458	3	US-09-457-046B-61		Sequence 61, Appl
18	176.5	7.4	306	3	US-09-457-046B-22		Sequence 22, Appl
19	175	7.4	455	3	US-09-457-046B-72		Sequence 72, Appl
20	173.5	7.3	460	3	US-09-457-046B-71		Sequence 71, Appl
21	157	6.6	438	3	US-09-457-046B-58		Sequence 58, Appl
22	154.5	6.5	450	3	US-09-457-046B-67		Sequence 67, Appl
23	150	6.3	303	3	US-09-457-046B-20		Sequence 20, Appl
24	148.5	6.3	440	3	US-09-457-046B-45		Sequence 45, Appl
25	146.5	6.2	459	4	US-09-457-046B-28		Sequence 2, Appl
26	146.5	6.2	459	4	US-10-074-279-2		Sequence 2, Appl
27	144	6.1	439	3	US-09-457-046B-28		Sequence 28, Appl

Db 58 LESLSSTLLYVAVRGKGTDCVIECNDGIGYVETAFVHQLHQBESNMLLLV 117
QY 121 PIKFSMEALSDERYPLLGQVNVFD-SGIAIGVSVSHKLDGTDACFLKSWGAVRGC 179
Db 118 GLSGFLSET--ETPLAAIQLNFKCGGLVIGAOFNHIIIGDMFTMTFMMNSWAKACRVG 174
QY 180 RENIHSLSEAL--FPRDRLPEKYDQMEALWFAGKAVATRRFVGVKALS 232
Db 175 IKEVAHTFTGLAPLPSAKVINIPPPS-----FEGVKFVSKRFVFNENAIT 221
QY 233 SIQDEAKSE-----SVKPSRVHVTGLMKHLIAASALTSGLTSTLSIAAQA 284
Db 222 RLKKEATEEDGDDQKKRPSRDLVTAFLSKLSIEMDCAKKEQKS-RPSLWVHMVN 280
QY 285 LRTRMNMETVLDN-ATGNLFWMAQALLESHTTPEISDL-----KLCD- 326
Db 281 LRKRTKL--ALENDVSGNFFTVNAESKIT-VAEKITDLTSLGSAAGCEIISVAKYDDA 337
QY 327 --LVNLLNGSVKQCNQDYFFTFKCKEGSRMCEYLDLQRTMSSMEPAPDIYLFSSWTF 384
Db 338 EVESMVLNSVREF--YYEWKGEK-----NPLITSWCRF- 371
QY 385 NPL----DFGNGRTSWIGVAGIESASCKF--IILVPTQCGSGIEAWNLBEKAMLEQ 438
Db 372 -PLVEVDGVM-----GIPSLVDTTAVPFGLVILVMDERAPAGDGLAVRACLSEHDMIQFQ 424
QY 439 DPHFLALAS 447
Db 425 HHQLLSYVS 433

RESULT 5
US-09-457-046B-68
; Sequence 68, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacyclases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Catharanthus roseus
US-09-457-046B-68

Query Match 16.3%; Score 387.5; DB 3; Length 439;
Best Local Similarity 26.8%; Pred. No. 3e-33;
Matches 128; Conservative 83; Mismatches 196; Indels 71; Gaps 21;
QY 1 MEKIEVSIN----SKHTIKPSTSP--LQPKLTLQLDLPAPXVPIVFFVPIPTHDFN 54
Db 1 MESKISIVETETLTKLKPS-SPTQSLSRNLSYNQNIYQTCVSGVFFENPD-GIE 58
QY 55 LPQTLADLQALSETLLTYFLSGRVKNNLYIDFEGVPVLEARNVCMDDTDFLRKIE 114
Db 59 ISTIREQLQNSLSTVSYYPFAGVKVKNVYIHCNDGIEFVEVIRCRMDILKYE--- 115
QY 115 CLNEFVPIKPSMEALSDERYPLLG-----VQNVVD-SGIAIGVSVSHKLDGTDAC 167
Db 116 -----LRSYARDVLPRVTVGSEDTTAVQLSHFDGGLAVAFGSHKVDGGTAS 168
QY 168 FLKSW--GAVFEGGRENIIHPSLSAALLFPPRDLLPEKYDQMEALWFAGKAVATRRFV 225
Db 169 FKWMAASACVLSSSHVPTPLLVSDS--IFPQDNI---ICEQFP-----TSKNCVEKTFI 220
QY 226 FGKAISSIQDEAKESVFKPSRVHVTGLMK-HLIAASALTSGLTSTLSIAAQA 283
Db 221 FPEPAIEKLSKAVFEGIEKPRVEVLNFTLSRCATVAGKSAKNNCQSLPFFVLQAI 280

QY 284 NLRTMNMETVLDNATGNLFWMAQALLESHTTPEISDLKCLDLVNLNGSVKQCNQDYF 343
Db 281 NLRPILELP---QNSVGNL-----VSIYFSRTIKENDYLNKEYTKLVINELRKEKQIK 332
QY 344 ETFKKEGIGYRMCSYLDQRTMSSMEPAP----DIYLFSSWTFN-FNPLDFGNGRTSWIG 398
Db 333 NLSREKLYTYAQM--EFVKSLEKFDISNFLDIDAYLSDSWCRFPFYDVFDFGNGKPIV- 389
QY 399 VAGKIESASCKP-----IILVPTQCGS--GIEAWNLBEKAMLEQDPHFLLALAS 447
Db 390 -----CLFQPIKXNVVMYDPFGDDYGIEAIVSEFQEKMSAFEXNEQLLFVS 438
RESULT 6
US-09-457-046B-59
; Sequence 59, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacyclases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-59

Query Match 13.0%; Score 308; DB 3; Length 331;
Best Local Similarity 26.8%; Pred. No. 8.5e-25;
Matches 93; Conservative 70; Mismatches 144; Indels 40; Gaps 15;
QY 117 NEFVPIKPSMEALSDERYPLLGQVNVFD-SGIAIGVSVSHKLDGTDACFLKSWGAV 175
Db 11 NELNKLHPFEHVSVD--VPLT-VQLTFEGGGLGALGSLKCLTALSGLIFVNSWAF 67
QY 176 FRGCRENIHPSLSAALLFPPRDLLPEKYDQME-ALWPAKGVATRRFVGVKALSSI 234
Db 68 ARGQTDLEIITPSF-DLAKMPPPCD-----IENLNMATGITKENIVTRRFVFLRSVESL 120
QY 235 QDEAKESVFKPSRVHVTGLMKHLIAASALTSGLTSTLSIAAQAANLRTMNMETV 294
Db 121 RERPSGNKKIRATRVESVLSFVWSRFMAST---NHDDKTGKIYTLIHPVNLRRQADPD-I 176
QY 295 LDNATGNLFWMAQALLESHTTPEI-----SDLKCLDLVNLNGSVKQCNQDYFETFKGE 350
Db 177 PDNMFEN-----IMRFSVTVPMALINENDEKASLVDMQREERKIDAVYVKLIQ-RJ 228
QY 351 GYGRMCEYLDQRTMSSMEPAPDIYLFSSWTFNFFNPL---DFGNGRTSWIGVAGKIESAS 407
Db 229 NRG-----HLEFLNKQASGFVNGEIVSFSFTSLCKFPVVEADFGWGPLMVASA---RMSY 281
QY 408 CKFIILVPTQCGSGIEAWNLBEKAMLEQDPHFLLALAS--PKTIL 452
Db 282 KNLVAFIDTKGEGIEAWNLQNDMSRFEADELLRYVSSNPSVMV 328

RESULT 7
US-09-457-046B-64
; Sequence 64, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacyclases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1


```
; SEQ ID NO 64
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-64

Query Match      10.2%; Score 242; DB 3; Length 461;
Best Local Similarity 25.4%; Pred. No. 2.2e-17;
Matches 120; Conservative 76; Mismatches 169; Indels 108; Gaps 27;

Qy 4 LEVINSKHTIKPSTSPLOP-----YKLTLLDQLTPPAVPIVFFYPTDHDNLPQ 57
Db 15 IPVTINQQLVHPS-SPTPANQSPHSYLSNLDIIGARVFTSPSYVYPSTNNKRESF-- 71

Qy 58 TLADLQALSETLTYYPPLSGRVK---NNLYIDDF---BEGVPYLEARVNCMDTFLRLRK 112
Db 72 VLKELQALSELVLPYPLSLGRLEVENKLEVFGEQGLMVLSANSSMDLAD----- 125

Qy 113 IECLNEFVPIKPFSEAL-----SDRY-----PLIGVQVNVFD-SGIAIGVSVSHKLDIG 162
Db 126 ---LGDLTVNPAPWLPLIFRNPGGEAYKILEMPLLIAQVTFTCGGFSLGIRLCHCICDG 182

Qy 163 GTADCFKLSGAVPRGCRNIHPS-----LSEALLFPRDDLP-EKYVDQMEA 211
Db 183 FGAMQFGLSNAATK-TGKLIADPEVPWDRTPKRPDPMPKYPHEVLPPIEBSNLNLS 243

Qy 212 LWFAGKV-----ATRRPVFGVKAISIQDRAKSESVPKSRV---HAVTGFPLWKHLIA 262
Db 242 LMDT-KPLQKCYRISKFCQCRVKSIAQED-----PVLVCTEDDAMAHIWRSNV- 290

Qy 263 ASRALTSGITSTRLSIAAQVNLTRNMETVLDNATGNLFWMAQAILEL-----BLSH 314
Db 291 --KALDVKPLDYNLRLTF-SVNVIRLETLLKRGFGYGNVVCCLACMSSVESLINDSLK 347

Qy 315 TTPEISDKLCLDVLNLSGSKVQCNQGYFETFKGKGEGYGRMCYLDLFTWSSMEPADDI 374
Db 348 TTRLVQDARL-----RVSEDIYLR-----WDYVDVVKRP-KRLEFGGKL 385

Qy 375 YLFSSWNFP--FNPLDFGNGRTSNGWAGKTB-SASCKFIILVPTQCGSGIEA 424
Db 386 TI-TOWTRFEMETADFGWGKFPVY---AGPIDLRPTQVCVLLPQ---GGVES 431

RESULT 8
US-08-207-904-2
; Sequence 2, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Anther-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 448 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-08-207-904-2

Query Match      9.7%; Score 230; DB 1; Length 448;
Best Local Similarity 23.1%; Pred. No. 4.3e-16;
Matches 107; Conservative 76; Mismatches 193; Indels 88; Gaps 20;

Qy 4 IEVINSKHTIKPSTSPLOPYKLTLLDQLTPPAYVPIVFFYPTDHDNLPOT---LA 60
Db 1 MKVSLKHWHVVKPA-EATWNGTVSLSCDQTFVAVTHVPTIYYRFC-HDC-LPSTDNIK 57

Qy 61 DLROALSETLTYYPPLSGRVK---NNLYIDDFBEGVPYLEARVNCMDTFLRLRKISCL 116
Db 58 TLRSLSKALWHFPLSLGRLEWIAAGSRLELDCNAGSIVLMEATEAKLGLDGLDFSPSPDL 117

Qy 117 NEFVPIKPFSEALSDERYPLLGQVNVFD-SGIAIGVSVSHKLDIGTADCFKLSWGAV 175
Db 118 NSLFRVDYT---IPIDELPLFLVQLTKFCGGIALSFALSHAVVQGSALYFLTEWASL 174

Qy 176 FRCGRNIIHPSLEALLFPDRDDLPEKYVDQMEALWAGKVVATRRFVFGVKAISIQ 235
Db 175 ARG--EPLNGEPFHDRKFL---RAGEPPIAYTTEHLQFNPPPLLGQ-----SSSE 221

Qy 236 DEAKSESVPKPSRVHANTGFLWKHLIAASRALTS---GTTSTRLSIAA----- 280
Db 222 EKKNET--KGSMLK-----LTKHQVEMLRKKAQNGOGRSYTRYEVVTAHWRACKAR 274

Qy 281 ----QAVNLTRNMETVLDNATGNLFWMAQAILEL---SETTPEISDKLCLDVLNLSG 333
Db 275 GHKEFQPTNLCLICVNIENIMQPLPKSY-FGNAIVDVIANGVSGDITSRPLEYVARRVA 333

Qy 334 SVKQCNQGYFETFKGKGEGYGRMCYLDLFTQRTWSSMEPADDIYLFSSWNFP--NP----- 386
Db 334 AIKMTSDYANS-----TIDFLKNQEDLSKYQDTHAFRSKESGPFYGNPNCZVI 381

Qy 387 -----LDFGNGRTSNGWAGKIESASCKFIILVPTQCGSG 421
Db 382 SWISLPLGLDFGKGEIHSFGTHEYDGDG---VILPGKEG 422

RESULT 9
US-08-207-904-17
; Sequence 17, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Anther-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 448 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-207-904-17

Query Match          9.7%; Score 230; DB 1; Length 448;
Best Local Similarity 23.1%; Pred. No. 4.3e-16;
Matches 107; Conservative 76; Mismatches 133; Indels 58; Gaps 20;

QY 4 IEVINSKHTIKPSTSTPLQPYKLLDQLTTPPAYVIVFFPIPTHDFNLQRLS 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKVSLKHHVWVKA-EATWNGTWSLSECDQTFVTHVETIYYRFC-HDC-LPSTDNI 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 DLQALSETLTYYPPLSGRVK---NNLYIDDPEEGVPLEARVNCMTDFLRKIECL 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 TLRTSLKALVHFVPLSGRLRWIAGSLRLDCLNAGSIVLMEATEAKLLDGLDFSPD 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 NEFVPIKPPGMEALSDRYPLGVQNVVD-SGIAIGVSVSHKLDGGTADCFKLSGAV 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 NSLPPRVDT--IPIDELPLFLVQLTKFQGGIALSFAISHAVVDGQSALYFLTEWAS 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 FRGRENIIHPSLSEALLPPRPDLDEKYVDQMEALWPAKGVATRRFVGVKAIISIQ 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 ARG--EPKNEPFDHDKFL---RAGEPIAYTFEHLQFNPPPLLLGQ-----SSSE 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 DEAKSEVPKPSRVHVTGFLWKHLIAASRALTS---GTTSTRLSTAA-----280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 EERKNET--KGSMLK-----LTKQVEMLRKKAQNGQGRSYTRYEVVTAHWECAKAR 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 -----QAVNLTRMMETVLDNATGNLFWMAQAILEL--SHTTPEISDLKLCDLWNLNG 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 GHRFEQPTNLCICVNIRNMQPLPKSY-FGNAIVDVIANGVSGDITSRPLEYVARRVRA 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 334 SVKQCNDYFETFKGEGYGRMCBYLDFQRTWSSMEPADPIYLFSSWTNFF--NP-----386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 AIRWVTSDYANS-----TIDFLKNOEDLSKYQDIHAFRSKEGPPFYGNPNLGI 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 -----LDFGWRGTSWIGVACKIESASCKFIILVPTQCGSG 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 SWISLPLLLGLDFGKGLIHSFGTHYDGDG---VILPCKEGDG 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-457-046B-69
; Sequence 69, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Pacitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-69
```

```
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-69

Query Match          9.6%; Score 228.5; DB 3; Length 451;
Best Local Similarity 22.6%; Pred. No. 6.4e-16;
Matches 99; Conservative 70; Mismatches 182; Indels 87; Gaps 19;

QY 13 TIRESTSTPLQP-----YKLLDQLTTPPAYVIVFFPIPTHDFNLQRLS 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 TFRKSYTIVPAEPTWSGRFFLAEDWDQVGTIITHITLYFYDKPSESFQ-GNVVEILKTSLS 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 ETLTYYPPLSGRV---KNLYIDDPEEGVPLEARVNCMTDFLRKIECLNEFVPIK 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 RVLVHFYPMAGRLKMLPRGFELMNCNAEGVEFIAESEGLSDF-----KDFSTP 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 124 PFS--MEAIS-----DERYPLGVQNVVD-SGIAIGVSVSHKLDGGTADCFKLSGAV 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 EFENLMPQVNVYKNPIETIPLFLAQVTKFKCGGISLVNVSHAVVDGQSALHLSGWGLA 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 177 RGRENIIHPSLSEAL-----LPPR-----DDLPEKYVDQMEALWPAKGVATR 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 RGELETV-PFLDRKILWAGEPLFPFVSPKPFHKEFDQPPFLIGETDNVERKKKTIW 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 RFVFGVKAISIODAKSESVPK-----SRHAVTGLWKHLIAASRALTSQTTSTRLSI 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 MLPLSTSQLKIRSKANGSKHSDAKGFTRYETVTVGHWR--CAKARGHSPEQPTALGI 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 AAQAVNLTRMMETVLDNATGNLFWMAQAILEL--SHTTPEISDLKLCDLWNLNGSVK 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 C-----IDTRSRMBPPLPRG-----YFGNATLDVVAASTGELISNELGFAASLISKAIX 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 QCNGCDYFETFKGEGYGRMCBYLDFQRTWSSMEPADPIYLFSSWTNFF--NP-----386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 342 NVTNEYVMI-----GIEYLNKQDLKKFQ---DLHALGTEGPPFYGNPNLGVSWL 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 -----LDFGWRGTSWIG 398
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 TLEMYGLDFGKGEFFYTG 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-457-046B-73
; Sequence 73, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Pacitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-73

Query Match          9.5%; Score 225; DB 3; Length 445;
Best Local Similarity 22.7%; Pred. No. 1.5e-15;
Matches 106; Conservative 86; Mismatches 181; Indels 94; Gaps 25;

QY 4 IEVINSKHTIKPSTSTPLQPYKLLDQLTTPPAYVIVFFPIPTHDFNLQRLS 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSIQIKQSTWVRPA-EETPNKSLMLSNIDMLIRTPYSHTGAVLIYKQPDNEDNIHPSSS 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 ---DFNLQRLSSETLYYPLSGRVK---LYIDDFEGVPLEARVNCMT 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 MYFDANI-----LIEALSALVFPYPMAGRLKINGDRYEDCNAGALFVEAESHVLE 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 DFLRLRKIECLNE-FVPIKPSMEBAISDERYPLGVQNVVD-SGIAIGVSVSHKLDGG 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 114 DFGDFRDNELHVRVWPTCDYS-KGISS---FPLLMVQLTRFCGGVSGIFQAHHHVCQGM 170
Qy 164 TADCFELKSWGAVFEG-----CRENIH-----PSISEAALLFPP-RDDLPKYYDQM 209
Db 171 AHPEFNNSWARIKGLLPALPEVDRVHLHPRNPQIKYSHSOFEBFVPSLPELLED-- 228
Qy 210 EALWFAGKQVATR-FVFGVKALSIIODEAK-SBSVPKPRVHAVTGELMKHGLIAASRAL 267
Db 229 -----CKTNKSQTLFILSRQINTLQKLDLSNNTLSTLSTYEVAAHWRS-VSKARGL 281
Qy 268 TSGTTSRLSIAQAVNLRTMMETVLDNATGNLFWMAQAILELSHTTPEISDLK---L 324
Db 282 -SDHEEIKL---IMPVGRSRINPSPKGYCGN-----WFLAVCTATVGLSCNPL 330
Qy 325 CDLVNLNLSGVKCKGQYFFETFKGEGYRMCXYLDFQRTMSSM-----EPAPDIY--- 375
Db 331 TDTAGKQVQKALGLDDYLS-----AIDHTSKPGLPVPMYSGPEKTLFNV 378
Qy 376 LFGSWTNF-FNPLDFGWRGTSWIGVAGKIESASCKFIILVPTQCGSG 421
Db 379 LVNSWGRIPQAMDFGWSGPTFFGISNIFVDGQC---FLIPSRDGDG 422

RESULT 12
US-09-457-046B-65
; Sequence 65, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-65

Query Match 9.4%; Score 222; DB 3; Length 497;
Best Local Similarity 22.9%; Pred. No. 3.8e-15;
Matches 115; Conservative 86; Mismatches 159; Indels 142; Gaps 28;
Qy 6 VSINSKHTIKPSTSTPLQPKLTL-LDQLTTPPAYVPIVFFYPIITDHDNLPOTLA--D 61
Db 55 VTTRKQVI---TAALPLQDHWLPLSLNLDLLPLNTHVCECYKKPLFTN---TVAYET 108
Qy 62 LQALSETLTYLPLSGRVKN-----LYDDFEGVPLVLEARNVDMTDFLELRKIEC 115
Db 109 LKTALETLVSYTAFAGELVTNPTGEPIILCNN--RGVDFVEAGADVLELRNLYDPDES 166
Qy 116 LNEFVPIKPSMEALSDERYPLLGVOVNVDSG-IAIGSVSHKLIIDGTDACFLKSWGA 174
Db 167 IAKLVPIK-----KHGVTAIQVTLQKGSIVVGCFTDHRVADAYSMNNFLLSWAE 216
Qy 175 VFR-----GC-----RENIHPSLSEAL-----LFPPRDDL-PEKYVDOMEAL 212
Db 217 ISKSDVPSICVPSRSLNPRRLVMDPSIDQIYMPVTSIPLPQETTNENL----- 269
Qy 213 WFAGKQVATRRFVFGVKALSIIODEAKSBSVPKSRVHAVTGFLWKHLIAASRALTSGTT 272
Db 270 -----LASRTIYIKANALQBLQTLASSKNGKRTKLESFAFLWK-LVAEHAARDVPPI 322
Qy 273 ST-RLSIAAQAQVNLRTMMETVLDNATGNLF---WMAQAILE-----LSHTTPEISDLK 323
Db 323 KTSKLGIT---AVDGRRL-MEKENVTFGNVLSVFPQGRIDDLISPLSWVTEVH--- 375
Qy 324 LCDLVNLNLSGVKCKGQYFFETFKGEGYRMCXYLDFQRTMSSMEPAPDI--- 374
Db 376 -----RFLKKSVT-----KEHFLNLDWVE-----TCRPTPAVSRIYSVSGSD 412

Qy 375 -----YLFSSWTFN-FNPLDFCWG-----RTSWIGVAGKIESASCKFIILVPTQCGSGI 422
Db 413 DGPAFVYSSGRSPVNVQDFGWSGSPVFGSYHFPGSSAG-----YVMDMP-----SSV 460
Qy 423 E-----AWVNLEBEKAMLEQD 439
Db 461 DDRDMVYLHLTKGLRFIEE 482

RESULT 13
US-09-457-046B-63
; Sequence 63, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-457-046B-63

Query Match 9.2%; Score 219; DB 3; Length 482;
Best Local Similarity 22.5%; Pred. No. 7.7e-15;
Matches 122; Conservative 93; Mismatches 171; Indels 156; Gaps 26;
Qy 1 MEKIEVSINSKHTIKPSTSTPLQPKLTLDDQLTTPPAYVPIVFFYPI-----TDHDFN 54
Db 4 LEX-SVTIISGRNVFPDQKST-LVDLKLVSVD-----LPMLSCHVYIQKGLCTCPNLP 54
Qy 55 LPQTLADLQALSETLTYLPLSGRV---KNNLVYDDFEGVPLEAR-----V 100
Db 55 LPALISHLKLSLITLTHFPPLAGRLSTSSGHVFLTCNDAGDFVFAQAKSHVSDVIA 114
Qy 101 NCDMTDFLRLRKIECLNEFVPIKPSME-AISDERY--PLIGVOVNVFDSGIALGVSVSH 157
Db 115 GIDVPDVVK-----EF-----FTYDRAVSVEGHRPILAVQVTELDNGVFIGCSVNH 161
Qy 158 KLIDGGTADCFKLSWGAVFRCGRENIHPHLS-FAALLFPPRDDLPKYYVDOMEALWPAQ 216
Db 162 AVTDGTSLWNFINPTFAEVSRAKNTVTRQPDFTRESVLISPAVLKVPQ-----GG 210
Qy 217 KKYA-----TERFVFGVKAISIQ-----DE--- 237
Db 211 PKVTFDENAPLRERKIFSRFSIQELKAVNVNKKMLTVNDGEIDGVLLGQSNKLNKG 270
Qy 238 -----AKSESVPKP-----SRVHAVTGFLWKHLIAASRALTSGTTSTRLSIA 279
Db 271 ENGLITLMLSLFGRNDVAVSKPVAVEISSFQSCALLWRAITRAKLPSSKXTTFRM--- 327
Qy 280 AQAVNLTRNME---TVLDNATGNLFWMAQAILELSHTTPEISDLKLC-DLVNL----- 330
Db 328 --AVNCHRLSPKLNPEYFGNAIQSVPTFAAABWLSR-----DLKWCADQLNQSAAH 379
Qy 331 LNSGVKCKGQYFFETFKGEGYRMCXYLDFQRTMSSMEPAPDIYLFSSWTFNFPNPLDFG 390
Db 380 ODGIRSVWADM-----EANPRCFPLGNADGASVTMGSSPRFPMDN-----DFG 424
Qy 391 WGRTSWT--GVAGKIESASCKFIILVPTQCGSG-IEAVVNLEBEKAMLEQDPHFLALAS 447
Db 425 WGRPVAVRSGRSNKFDCKISAF----FGREGNGTVDLVLSLPSWTAGIESDGEFMRVYT 480
Qy 448 PK 449
Db 481 NK 482

RESULT 14

US-09-457-046B-50
 ; Sequence 50, Application US/09457046B
 ; Patent No. 6287835

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679

; CURRENT APPLICATION NUMBER: US/09/457,046B

; CURRENT FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 50

; LENGTH: 443

; TYPE: PRT

; ORGANISM: Taxus cuspidata

US-09-457-046B-50

Query Match

7.5%; Score 179; DB 3; Length 443;

Best Local Similarity 23.8%; Pred. No. 1.5e-10;

Matches 101; Conservative 64; Mismatches 176; Indels 84; Gaps 21;

QY 1 MEK-----LEVSINSHKTIKPTSTPLQYKLTLLDQTPP-AVPIVFFYPITDHDEN 54

DB 1 MEKSSADLHNIIRVVVAP-CQTPKTIQLQSSIDKMGGFANVLLVFG---ASHGV 56

QY 55 LPQTLD---LRQALSETLTYPLSGRVK---NNLYIDDPEGVPLRARNVNDMTD 106

DB 57 -----ADPAKTRRALSKTLYVFFPAGRLKKEDGIEVEGIEQCALFVEMADND--- 108

QY 107 FLRLKIECLNEFVPIKPFSEMAIS-DEHYPP---LLGVQNVFD-SGIAIGVSVSHKLD 161

DB 109 ---LSWRDLQENLFRQLQSSLDTDYKDLHLMTVQVTPFTCGGFMVGTSHVQSICD 165

QY 162 GGTAPCFELKSWGAVRGCRNIIHPSLSEALFLPPRDLP-EKYVDQMEALWFAGKVA 220

DB 166 GNLGCFKFKMAEIVRGEVKEPIEWNRE--LVKPEDVIHLQLYSEIRPLVVEKVG 223

QY 221 TRRFVFGVKAISIOEAKSESVKPSRVHVTGFLWKHLIAASRALTSGTSTRLSIAA 280

DB 224 QTSLVISPEKINHAKCIEESESSESPFIVTAMVW---LARTAFQIPHN----- 272

QY 281 QAVNLRTRMMETVLD-----NATGNLFWMAQAILELSTTPEISDLKCLDLVNLINGS- 334

DB 273 EDVTLILAMDARRSPDPIPKGYGNVI-----GTYAKD-----NVNLLSGSL 317

QY 335 -----VKOCNGDYPTFKGEGYGRMCEYLDQFTWSSMEPAPDIYLFSSWTFN-FNPL 387

DB 318 LHALTVIKSMSSFYENMTSR-----VLNPFSTLDLSMKYENVVLSLSDWSRLGHNEV 369

QY 388 DFGWG 392

DB 370 DFGWG 374

RESULT 15

US-09-457-046B-70

; Sequence 70, Application US/09457046B

; Patent No. 6287835

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

; FILE REFERENCE: 53679

; CURRENT APPLICATION NUMBER: US/09/457,046B

; CURRENT FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 74

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 70

; LENGTH: 461

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-457-046B-70

Query Match

7.5%; Score 178.5; DB 3; Length 461;

Best Local Similarity 24.3%; Pred. No. 1.8e-10;

Matches 118; Conservative 68; Mismatches 197; Indels 103; Gaps 26;

QY 1 MEKIEVSINSKHTIKPSTSTPLQYKLTLLDQTPP-AVPIVFFYPITDHDENLQTLA 60

DB 13 LEKKEVEL-----VKPS-KHTHCETLSLSLQDN-DPFNEVMYATIVYFVANGKNDOPVS 65

QY 61 DLROALSETLTYPLSGRV-----NNLYIDDPEGVPLRARNVNDMTDFLRKIKIC 115

DB 56 LLRKALSELVHYPLSGKLMSESGKQLQVYLGEVGFVETATSTLDSS---LNYIEN 122

QY 116 LNEFVPIK---PFSMEAISDERYPPLLGQVNVFD-SGIAIGVSVSHKLDG-GTADCFLK 170

DB 123 LDDQVALRLVPEIDIDYESNVCHPLALQVTFACGGFTIGTALTATHAVCDGKVA----- 177

QY 171 SWGAVRGRGCRNIIHPSLSEAL-----LPPRDDLPLEK-----YVDQMEA 211

DB 178 -----QIIH-ALTELAAGKTEPSVKSWQORERLVGKIDNKPQKVPQSHIDGFLA 225

QY 212 --LWFAGKKVATRRFVFGVKAISLQD-----EAKSESVKPSRVHVTGFLWKHLIAA 263

DB 226 TSAYLFTTDDVVTETINIRAGDIXRLKDSMMKECEYLKESF---TTEVLSSYIWK---LR 279

QY 264 SRALTSGTSTTRLSIAAQAIVNLRTRMMETVLD-----NATGNLFWMAQAILELSTTPE 318

DB 283 SRALKLNPDG--ITVLGVAWGIR-----HVLDPPLPKGYGNAY-IDVYVEL--TVRE 327

QY 319 ISDLKCLDLVNLINGSVKQC--NGDYPTFKGEGYGRMCEYLDQFTWSSMEPAPD-IY 375

DB 328 LBESSISNIANEVKKAKTAYBKGYIEELKNT-----RLMRDSDSMFEGVSDGLF 378

QY 376 LFSSWTN--FPNPLDQNGRTSWIGVAGKIESASCKFIILVPTQCGSGTEAWN--LEEE 431

DB 379 FUTDNRNIGWFGSMDPGWNEPVLRLPTORESTVHVGMILKPSKSDPSMEGGVAVIKLP 438

QY 432 KWAMLE 437

DB 439 RDAMVE 444

Search completed: July 6, 2004, 13:40:15

Job time : 20.1722 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:38:51 ; Search time 49.964 Seconds

(without alignments)
2816.033 Million cell updates/sec

Title: US-09-857-518A-6

Perfect score: 2373
Sequence: 1 MEKIEVINSKHTKPTSS.....MAMLEQDEHFLALASPXTLI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	21.0	436	9	US-09-866-572A-62
2	498	21.0	436	9	US-09-866-570A-62
3	498	21.0	436	14	US-10-167-603C-9
4	498	21.0	436	15	US-10-166-984-66
5	485	20.4	446	9	US-09-866-572A-62
6	485	20.4	446	9	US-09-866-570A-62
7	485	20.4	446	14	US-10-166-984-66
8	485	20.4	446	15	US-10-166-984-66
9	448	18.9	435	9	US-09-866-572A-62
10	448	18.9	435	9	US-09-866-570A-62
11	448	18.9	435	14	US-10-166-984-66
12	448	18.9	435	15	US-10-166-984-66
13	438	18.5	474	12	US-10-167-603C-9
14	438	18.5	474	12	US-10-167-603C-9
15	438	18.5	474	12	US-10-167-603C-16

16	404.5	17.0	433	9	US-09-866-572A-66
17	404.5	17.0	433	9	US-09-866-570A-66
18	404.5	17.0	433	12	US-10-167-603C-9
19	404.5	17.0	433	14	US-10-166-984-66
20	404.5	17.0	433	15	US-10-166-984-66
21	404.5	17.0	433	15	US-10-166-984-66
22	387.5	16.3	439	9	US-09-866-572A-68
23	387.5	16.3	439	9	US-09-866-570A-68
24	387.5	16.3	439	12	US-10-167-603C-8
25	387.5	16.3	439	14	US-10-166-984-68
26	387.5	16.3	439	15	US-10-166-984-68
27	386	16.3	457	14	US-10-259-549-2
28	386	16.3	460	14	US-10-259-549-7
29	385	16.2	408	14	US-10-259-549-15
30	377	15.9	406	14	US-10-259-549-17
31	374	15.8	403	14	US-10-259-549-23
32	357.5	15.1	331	9	US-09-866-572A-59
33	308	13.0	331	9	US-09-866-570A-59
34	308	13.0	331	14	US-10-166-984-61
35	308	13.0	331	15	US-10-166-984-61
36	307	12.9	251	12	US-10-424-599-152518
37	266.5	11.2	152	12	US-10-424-599-257533
38	259.5	10.9	449	16	US-10-437-963-108868
39	253	10.7	448	12	US-10-424-599-222595
40	253	10.7	449	12	US-10-425-114-56357
41	249.5	10.5	446	16	US-10-437-963-178429
42	249.5	10.5	467	12	US-10-424-599-237005
43	247	10.4	468	12	US-10-425-114-42259
44	247	10.4	475	12	US-10-425-114-47544
45	247	10.4	475	12	US-10-425-114-47544

ALIGNMENTS

US-09-866-572A-62

Sequence 62, Application US/09866572A

Patent No. US20020138859A1

GENERAL INFORMATION:

APPLICANT: Croteau, Rodney et al.

TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway

FILE REFERENCE: 53679

CURRENT APPLICATION NUMBER: US/09/866,572A

CURRENT FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: 09/457,046

PRIOR FILING DATE: 1999-12-07

NUMBER OF SEQ ID NOS: 74

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 62

LENGTH: 436

TYPE: ERT

ORGANISM: Arabidopsis thaliana

US-09-866-572A-62

Query Match 21.0%; Score 498; DB 9; Length 436;

Best Local Similarity 30.9%; Pred. No. 8.9e-44;

Matches 145; Conservative 90; Mismatches 184; Indels 50; Gaps 19;

QY	1	MEKIEVINSKHTKPTSS	---	LQPKLILDLQTPPAVIVFFPITDHD	---	PNL 55
DB	1	MEK-NVELISREIVKPS	-SPTPDDKRIINLSILDLSSPMYTGALLFYAADPQNLLGFST	58		
QY	56	PQTALDLQALSETLTYPLSGRVKNLYIDDDPEGVPYLEARNVCMDFLRKTEC	115			
DB	59	EETSLKLLKSLSKTLPFYPLAGRIIGS	-FVCEDEGAVFIARVDFHLLSEFLKCPVPS	117		
QY	116	LNEFFYKPFPMSEALSDERYPLLVQVNVFO	-SGTAVGVSVSHKLDIGTADCFILKSWGA	174		
DB	118	LELLIPVEAKSREAVT	---WPVLLIQANFFSCGLVITICVSHKITDALSAMPFGWAE	174		
QY	175	VFRGCRENIHPSLSEALFFPRDLPEKYVDQMEALWFAGKCVATRFVFGVAISST	234			

TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-166-984-75

Query Match 18.9%; Score 448; DB 14; Length 435;
Best Local Similarity 29.8%; Pred. No. 2e-38;
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;

3 KIEVINSKHITKSTSTPIQPKYKLTLLDQTPPAYVPIVFFPIHDDHNLPTLADL 62
4 KLEVT--GKEVIKPA-SPSPDRQLQSLDLYCPIYVSTIFFYDLITSESEV--FSEN 58
63 ROALSETLTYPLSGRVKNNLYIDDFEGVYPLEARVNCMDTDFLRKKEICLNEFVPI 122
59 KLSLSETLSRFPYLAGRIE-GLSISCNDEGAVFTEARTDILLPDLFLNLTDSLSGFLPT 117
123 KPFSMEALSDEYRPLLGQVNVF--DSGIAIGVSVSHKLIIDGGTADCFKKGAVFRGR 180
118 ---LAAGESPAWPLLSVKYTFFGSGGVAVSVSHKICDIAASLVTFVKWATTAKG 174
181 EN-----IHPSLSEALLPFRD---DLPEKYVDQMEALWFAKKVATRFVGV 228
175 SNSTIEFAATTIYPPPHMVEQFPSTDSNITSKYV-----LKRFPVEP 220
229 KAISIOEAKSEVPKPSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAAQAVALRTR 288
221 SKIAELKHAASESVVPVTRVEAIVSLIWR---CARNSRSNLLIPROAVMWAQANDIRL 277
289 MNMTVLDNATGNFWMAQAILELSTHTPEISDLKLCDLVNLNGSVKOCNGDYFTFKG 348
278 IPSSVAPKDVIGNL---QSGFSLKDAE--SEFEIPEIVATPRKNKERNEMIKESLQ 331
349 KEGYGRMCYLDPORTMSMEPADYILFSSWTNF-FNPLDFGNGRTSMIGVAGKIESAS 407
332 NT-IGQSLLSLMAETVSESTE--IDRYIMSSWCRKPFYEVDFGSGSPVWVGASHYIYN 388
408 CKFIILVPTCCSGGIEAWNLBEKMWALEQDPHFLALA 446
389 MGVVWLIDSKGDEGVAEWISLPEEDMSVFVDDQELLAYA 427

RESULT 12
US-10-166-984-75
Sequence 75, Application US/10166984
Publication No. US2004000562A9
GENERAL INFORMATION:

APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
FILE REFERENCE: 4630-62267
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: US 09/866,570
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 09/457,046
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 09/411,145
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 76
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 75
LENGTH: 435
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-166-984-75

Query Match 18.9%; Score 448; DB 15; Length 435;
Best Local Similarity 29.8%; Pred. No. 2e-38;
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;

3 KIEVINSKHITKSTSTPIQPKYKLTLLDQTPPAYVPIVFFPIHDDHNLPTLADL 62
4 KLEVT--GKEVIKPA-SPSPDRQLQSLDLYCPIYVSTIFFYDLITSESEV--FSEN 58

63 ROALSETLTYPLSGRVKNNLYIDDFEGVYPLEARVNCMDTDFLRKKEICLNEFVPI 122
59 KLSLSETLSRFPYLAGRIE-GLSISCNDEGAVFTEARTDILLPDLFLNLTDSLSGFLPT 117
123 KPFSMEALSDEYRPLLGQVNVF--DSGIAIGVSVSHKLIIDGGTADCFKKGAVFRGR 180
118 ---LAAGESPAWPLLSVKYTFFGSGGVAVSVSHKICDIAASLVTFVKWATTAKG 174
181 EN-----IHPSLSEALLPFRD---DLPEKYVDQMEALWFAKKVATRFVGV 228
175 SNSTIEFAATTIYPPPHMVEQFPSTDSNITSKYV-----LKRFPVEP 220
229 KAISIOEAKSEVPKPSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAAQAVALRTR 288
221 SKIAELKHAASESVVPVTRVEAIVSLIWR---CARNSRSNLLIPROAVMWAQANDIRL 277
289 MNMTVLDNATGNFWMAQAILELSTHTPEISDLKLCDLVNLNGSVKOCNGDYFTFKG 348
278 IPSSVAPKDVIGNL---QSGFSLKDAE--SEFEIPEIVATPRKNKERNEMIKESLQ 331
349 KEGYGRMCYLDPORTMSMEPADYILFSSWTNF-FNPLDFGNGRTSMIGVAGKIESAS 407
332 NT-IGQSLLSLMAETVSESTE--IDRYIMSSWCRKPFYEVDFGSGSPVWVGASHYIYN 388
408 CKFIILVPTCCSGGIEAWNLBEKMWALEQDPHFLALA 446
389 MGVVWLIDSKGDEGVAEWISLPEEDMSVFVDDQELLAYA 427

RESULT 13
US-10-167-603C-7
Sequence 7, Application US/10167603C
Publication No. US20030226161A1
GENERAL INFORMATION:

APPLICANT: ZENK, Meinhard H.
APPLICANT: KITCHAN, Toni
APPLICANT: GROTHE, Torsten
TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives
FILE REFERENCE: 67529
CURRENT FILING DATE: 2002-06-11
PRIOR APPLICATION NUMBER: EP 0114122.3
PRIOR FILING DATE: 2001-06-11
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 7
LENGTH: 474
TYPE: PRT
ORGANISM: Papaver somniferum
US-10-167-603C-7

Query Match 18.5%; Score 438; DB 12; Length 474;
Best Local Similarity 27.9%; Pred. No. 2.6e-37;
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

1 MEKIEVINSKHITKSTSTPIQPKYKLTLLDQTPPAYVPIVFFPIHDDHNLPTLADL 49
4 MYSAAVEVISEKITKPTTPTSQLKNFNLSDQCPDYLYYVPIILFPATAANSTGSSN 63
50 DHDNLPQTADLROALSETLTYPLSGRVKNNLYIDDFEGVYPLEARVNCMDTDFLR 109
64 HHD-----DLDLKSSLSKTLVHFYPMAGRMIDNLDVCHDQGINFYKVKIRGKCEMS 118
110 LRKIECLNEFVPIKPSMEALSDEYRPLLGQVNVF--SGIAIGVSVSHKLIIDGGTADCF 168
119 QPDVP-LSQLPSEVWS---ASVPEALVIVQVNMEDCGGTACSSVSHKIAADATMTTF 174
169 LKSGAVFRGR-----ENIHPSLSEALLPFRD-----PEKY 205
175 IRWSASTTKTSRSGSGSTAATVDQKLIPSF-DSASLFPSPERLTSPSGMSETPFSPSTPDT 233
206 VDQMEALWFAKKVATRFVGVKAISIQ-----DEAKSEVPKPSRVHVTGFLW 257

Db 234 EDD-----KTVSKRFVDFPAKITSVREKLOVLMDHNYKR---RQTRVVFVTSLIW 281
QY 258 KHLIAASRALTSCTTSTRELSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAPLPAITTTT 333
QY 317 -----PEISDLKLCOLVNLINGSVKQCNDYFETFKGEGYGRMCCEYLDFO 362
Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKDGKSLKVIQNFASGHD----- 388
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNLDGFGRTSWIGVAGKIE-SASCKFIILVPT 416
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWGPFIWTVDPNIRKPKNCFF--MNDT 443
QY 417 QCGSGIEAWNLLEBKMALE 437
Db 444 KCGEGIEVWASFLDDMAKFE 464

RESULT 14
US-10-167-603C-14
; Sequence 14, Application US/10167603C
; Publication No. US20030226161A1
; GENERAL INFORMATION:
; APPLICANT: ZENK, Meinhard H.
; APPLICANT: KUTCHAN, Toni
; APPLICANT: GROTHE, Torsten
; TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 67529
; CURRENT APPLICATION NUMBER: US/10/167,603C
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: EP 01114122.3
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 14
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Papaver somniferum
US-10-167-603C-14

Query Match 18.5%; Score 438; DB 12; Length 474;
Best Local Similarity 27.9%; Pred. No. 2.6e-37;
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

QY 1 MEKIEVSINSKHTIKPST-SSTPLQPYKLTLLDQTP-PAYVPIVFFVPIIT----- 49
Db 4 MYSAAVEVISKETIKPTTTPSQKNFNLSDQCFPLYVYVPIILFPATAANSTGSSN 63
QY 50 DHDENLPTQLADLQALSETLTYPLSGRVKNLYIDDFEEGVPILEARNVCMDFLR 109
Db 64 HED-----DLDLKSSLSKTLVHVEYPMAGRMIDNILDCHDQGINFYKVKIRKMCCEMS 118
QY 110 LRKIECLNEFVPIKPFSEALSDERYPLLGVOVNVFD-SGIAIGVSVSHKLIIDGTTACF 168
Db 119 QPDVP-LSQLLPSEWVS---ASVPKEALVIVQVNMFDGCGTAICSSVSHKIADAATMSTF 174
QY 169 LKSGAVRGR-----ENIHPSLSAALLFPDRL-----PEKY 205
Db 175 IRSWASTTKTSRSGSTAANVDQKLIPIF-DSASLFPSERLTSPSGMSEIPFSTPDT 233
QY 206 VDQMEALWFAGKVKVATRRFVGVKALSIQ-----DEAKSEVPKPSRVHVAVTGLW 257
Db 234 EDD-----KTVSKRFVDFPAKITSVREKLOVLMDHNYKR---RQTRVVFVTSLIW 281
QY 258 KHLIAASRALTSCTTSTRELSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAPLPAITTTT 333
QY 317 -----PEISDLKLCOLVNLINGSVKQCNDYFETFKGEGYGRMCCEYLDFO 362

Db 234 EDD-----KTVSKRFVDFPAKITSVREKLOVLMDHNYKR---RQTRVVFVTSLIW 281
QY 258 KHLIAASRALTSCTTSTRELSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAPLPAITTTT 333
QY 317 -----PEISDLKLCOLVNLINGSVKQCNDYFETFKGEGYGRMCCEYLDFO 362

Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKDGKSLKVIQNFASGHD----- 388
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNLDGFGRTSWIGVAGKIE-SASCKFIILVPT 416
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWGPFIWTVDPNIRKPKNCFF--MNDT 443
QY 417 QCGSGIEAWNLLEBKMALE 437
Db 444 KCGEGIEVWASFLDDMAKFE 464

RESULT 15
US-10-167-603C-16
; Sequence 16, Application US/10167603C
; Publication No. US20030226161A1
; GENERAL INFORMATION:
; APPLICANT: ZENK, Meinhard H.
; APPLICANT: KUTCHAN, Toni
; APPLICANT: GROTHE, Torsten
; TITLE OF INVENTION: Salutaridinol 7-O-Acetyltransferase and derivatives
; TITLE OF INVENTION: thereof
; FILE REFERENCE: 67529
; CURRENT APPLICATION NUMBER: US/10/167,603C
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: EP 01114122.3
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 16
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Papaver somniferum
US-10-167-603C-16

Query Match 18.5%; Score 438; DB 12; Length 474;
Best Local Similarity 27.9%; Pred. No. 2.6e-37;
Matches 140; Conservative 87; Mismatches 170; Indels 104; Gaps 21;

QY 1 MEKIEVSINSKHTIKPST-SSTPLQPYKLTLLDQTP-PAYVPIVFFVPIIT----- 49
Db 4 MYSAAVEVISKETIKPTTTPSQKNFNLSDQCFPLYVYVPIILFPATAANSTGSSN 63
QY 50 DHDENLPTQLADLQALSETLTYPLSGRVKNLYIDDFEEGVPILEARNVCMDFLR 109
Db 64 HED-----DLDLKSSLSKTLVHVEYPMAGRMIDNILDCHDQGINFYKVKIRKMCCEMS 118
QY 110 LRKIECLNEFVPIKPFSEALSDERYPLLGVOVNVFD-SGIAIGVSVSHKLIIDGTTACF 168
Db 119 QPDVP-LSQLLPSEWVS---ASVPKEALVIVQVNMFDGCGTAICSSVSHKIADAATMSTF 174
QY 169 LKSGAVRGR-----ENIHPSLSAALLFPDRL-----PEKY 205
Db 175 IRSWASTTKTSRSGSTAANVDQKLIPIF-DSASLFPSERLTSPSGMSEIPFSTPDT 233
QY 206 VDQMEALWFAGKVKVATRRFVGVKALSIQ-----DEAKSEVPKPSRVHVAVTGLW 257
Db 234 EDD-----KTVSKRFVDFPAKITSVREKLOVLMDHNYKR---RQTRVVFVTSLIW 281
QY 258 KHLIAASRALTSCTTSTRELSIAAQAQVNLRTMMNMTVDNATGNLFWMAQAILLSHTT- 316
Db 282 K-----SVMKSTPAGFLPVVHVAVNLKMKD-PPQDVSVFGLSVTVSAPLPAITTTT 333
QY 317 -----PEISDLKLCOLVNLINGSVKQCNDYFETFKGEGYGRMCCEYLDFO 362
Db 334 TNAVNTINSTSESQVVLHELHDFIAQMRSEIDKVKDGKSLKVIQNFASGHD----- 388
QY 363 RTWSSMEPADDI-----YLFSSWTFNFPNLDGFGRTSWIGVAGKIE-SASCKFIILVPT 416
Db 389 ---ASIKKINDVEINFWISSWCRMGLEYIDFGWGPFIWTVDPNIRKPKNCFF--MNDT 443
QY 417 QCGSGIEAWNLLEBKMALE 437
Db 444 KCGEGIEVWASFLDDMAKFE 464

Tue Jul 6 14:13:41 2004

us-09-857-518a-6.rapb

Page 8

Search completed: July 6, 2004, 13:48:41
Job time : 50.964 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:50 ; Search time 19.1722 Seconds
(without alignments)
2267.788 Million cell updates/sec

Title: US-09-857-518A-6
Perfect score: 2373
Sequence: 1 MEKIEVSINSKTIKPTSS.....MAMLEQDPHELALASPKTLI 452

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191536 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:
1: Pirl:
2: Pirl:
3: Pirl:
4: Pirl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	21.0	436	D86378	protein F21J9.8 [i
2	485	20.4	446	D71418	hypothetical prote
3	448	18.9	435	E71418	hypothetical prote
4	262	11.0	430	T46216	hypothetical prote
5	242	10.2	446	T10711	anthranilate N-ben
6	242	10.2	461	T00918	hypothetical prote
7	234.5	9.5	461	T51458	hypothetical prote
8	233.5	9.8	445	T10719	anthranilate N-ben
9	232	9.8	442	T10718	anthranilate N-ben
10	228.5	9.6	451	T00527	hypothetical prote
11	228	9.6	464	G86453	YUP8H12R.39. homol
12	225	9.5	445	T10717	anthranilate N-ben
13	222	9.4	572	T01056	hypothetical prote
14	219	9.2	452	G84823	probable anthocyan
15	217	9.1	454	T45611	N-hydroxycinnamoyl
16	215	9.1	451	T01140	hypothetical prote
17	209	8.8	456	T45610	proanthranilate N-
18	207	8.7	464	T45612	N-hydroxycinnamoyl
19	200.5	8.4	450	T45573	anthranilate N-hyd
20	196	8.3	443	T45574	anthranilate N-hyd
21	184	7.8	459	T45576	hypothetical prote
22	180	7.6	450	T48479	hypothetical prote
23	178.5	7.5	461	H84644	hypothetical prote
24	177.5	7.5	458	T06313	hypothetical prote
25	177	7.5	421	T05583	CER2 protein Ara
26	175	7.4	455	T09666	probable anthranil
27	173.5	7.3	460	T03274	hser201 protein, hy
28	172	7.2	428	T48008	hypothetical prote
29	166.5	7.0	436	T00908	hypothetical prote

protein F21J9.8 [i
hypothetical prote
hypothetical prote
protein F1K23.12 [i
anthranilate N-ben
hypothetical prote
hypothetical prote
10-deacetylribacat
taxadienol acetyl
hypersensitivity r
hypothetical prote
hypothetical prote
alanine-tRNA ligas
MAPK delta-1 prot
trichothecene 3-O-
bacitracin synthe

30 163 6.9 465 2 D86166
31 159.5 6.7 648 2 T45575
32 159 6.7 433 2 H84826
33 158.5 6.7 480 2 H84411
34 157 6.6 475 2 T45961
35 154.5 6.5 450 2 T02368
36 154 6.5 469 2 C86170
37 148.5 6.3 440 2 T52320
38 144 6.1 439 2 T52321
39 142 6.0 447 2 T45653
40 139 5.9 426 2 S60660
41 132.5 5.6 414 2 T13436
42 102.5 4.3 900 2 S73748
43 101 4.3 886 2 T48544
44 100 4.2 451 2 T43733
45 99.5 4.2 2607 2 T31678

ALIGNMENTS

RESULT 1
D86378 protein F21J9.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86378
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alorco
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86378
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-436 <STO>
A:Cross-references: GB:AB005172; NID:g9743331; PIDN:AAF97955.1; GSPDB:GN00141
C:Genetics:
A:Gene: F21J9.8
A:Map position: 1

Query Match 21.0%; Score 498; DB 2; Length 436;
Best Local Similarity 30.9%; Pred. No. 9.5e-34;
Matches 145; Conservative 90; Mismatches 194; Indels 50; Gaps 19;

QY 1 MEKIEVSINSKTIKPTSSSTP--LQPYKTLIDQLTPPAYVPVFPYPTDHD---FNL 55
Db 1 MEK-NVEILSREIVKFS-SPTDDKRIILSLDILSSPMYTGALLEAYADPQLLGST 58
QY 56 PQTADLRQALSETLTYLYPELGRVKNLYIDFEEGVPYLEARVNCMDTDFRLRKIEC 115
Db 59 EETSLLKLSKLSKTLPIFYPLAGRIIGS-FVECNDEGAVFEARVDHLLSEFLKCPVES 117
QY 116 LNFVPIKIPSPMAISDERYPILGVQVNFED-SGIALGVSVSHKLDGGTADCFLKSWGA 174
Db 118 LELLIPVEAKSREAVT---WPVLLIQANFFSCGLVITICVSHKIDTATSLAMFIRGAE 174
QY 175 VFSGRENITHPSLSRAALLFPREDLPKRYVQMEALWPAKGVATRRVFGVKAISSI 234
Db 175 SSRGLGITLI-PSFTASEVFPKPLDELPSKPMRKEV--SEMSCVTKRFVFDASKIKKL 231
QY 235 QDEAKSESVKPSRVHATVGLWKHLIAASRALTSCTTSTRLSIAAQAQVNLRTMNMETV 294
Db 232 RAKASRNLVKNPITVEAVTALFWRCVTKVSRL---SSLTPRTSVLQILMNLGRK--VDSL 286
QY 295 LDNATGNLFWN-----AQAILLSHTTPEISDLKCLDVLNLLSGSVKQCGNDYFETFK- 347

Db 287 CENTIGNLSMLKNEEAIERIQVDVDIRAK--EIFSLNCKBMSKSSSRIFELLE 344
QY 348 -GREGYGRMCEYLDLDFORTSMSEPPADYILFSSWTFN-FNPLDFGWRGTSMIGVAGKIES 405
Db 345 IGKV-YGRGNEM-----DLWMSNCKGLGYDADFWMGKPVW--VTGRGTS 387
QY 406 ASCKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFLALAS--PKTLI 452
Db 388 HFKNMLLIDTKGEGIEAMITLIEQMSLFECDQELLESASLNPPVLI 436
RESULT 2
E71418
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: D71418
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chaiwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: D71418
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-446 <BRV>
A:Cross-references: GB:297338; NID:g2244870; PID:g2244896
C:Genetics:
A:Map position: 4COP9-4G3845
C:Superfamily: Arabidopsis CER2 protein
Query Match 20.4%; Score 485; DB 2; Length 446;
Best Local Similarity 29.1%; Pred. No. 1.2e-32;
Matches 143; Conservative 91; Mismatches 165; Indels 92; Gaps 23;
QY 4 IEVSINSKHTIKSTSTP--LQPKLTLDLQTERAVPIVFFYPTDHD-FNLPTLA 60
Db 6 MKVETTSKEILKES-SPTNNLQTLQUSIDHILPPVITVAFLY--TKNDLISQKTS 62
QY 61 DLQALSETLTYPLSGRVKNLYIDDPEEGVYLEARY-NCMDTDLRLKTECLNEF 119
Db 63 KLKTSLETITKFPYLAGRI-TGVVDTDBGAIFDARVNNCPLEFLKCPDFALQQL 121
QY 120 VPIKPSMEALSIDERYELLGWVVF--SGIAGVSVSHKLDIGGTADCFKSGAVFRG 178
Db 122 LPDWDVNDPYVAATWELLVAKATYFCGGNAIGICITHKIADAASITFIRSWAATARG 181
QY 179 CREN---IHPSLSEALLPFRDDLEPKYVDQMEALWFAKVA-TRRFVFGVKAIS 234
Db 182 --ENDAAAMESPVFAGANFPANEAFLPADEQ-----AGKRSITKRFVFEASKVEDL 234
QY 235 QDEAKS-ESVPKPSRVHVTGFLWKHLIAASRALTSGTSTRLSIAAQVNLKTR 293
Db 235 RTAAASSETVDQPRVESVTALIKCFVASK-----TTTCDHKVLQLANLRSKI-PSL 288
QY 294 VLDNATGNLFWNA-----QALLESHTTPEISDLKCD----- 326
Db 289 LQESSIGNLAFSSVVLISIGRGEVKIEAVRDLRKKBELGTVLIDBGGSDSSSMIGSK 348
QY 327 LVNLLNGSVQKQXGYFTFKGEGYGRMCEYLDLDFORTSMSEPPADYILFSSWTFNFP 386
Db 349 LANLM-----LTNYSKL-----SYETHEP-----YTSSCKL--P 377
QY 387 L---DFNGRTSMIGVAGKIESASCKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFL 443
Db 378 LYEASFQWDSFVW--VGVNVPVLGNLAWLIDSKDGGIEAFVTLPEENMSSPQNPPELL 435
QY 444 ALA--SPKTLI 452

Db 436 AFATMNPVLV 446
RESULT 3
E71418
hypothetical protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
C:Accession: E71418
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chaiwatzis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: E71418
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-435 <BRV>
A:Cross-references: GB:297338; NID:g2244870; PID:e326326; PID:g2244897
C:Genetics:
A:Map position: 4COP9-4G3845
Query Match 18.9%; Score 448; DB 2; Length 435;
Best Local Similarity 29.8%; Pred. No. 1.5e-23;
Matches 137; Conservative 79; Mismatches 193; Indels 50; Gaps 15;
QY 3 KIEVSINSKHTIKSTSTP--LQPKLTLDLQTERAVPIVFFYPTDHD-FNLPTLA 62
Db 4 KLEVT--GKEVIKPA-SFSPDRDLQSLIDLYCGIYVSTIFFYDLITESEV--FSEN 58
QY 63 RQALSETLTYPLSGRVKNLYIDDPEEGVYLEARYNCMDTDLRLKTECLNEFVPI 122
Db 59 KLSLSETLSRFPYLAGRIE-GSISCNDEGAVFTEARTDLLLPDLRLMNTDLSLGF 117
QY 123 KPSMEALSIDERYELLGWVVF--DSGIAIGVSVSHKLDIGGTADCFKSGAVFRG 180
Db 118 ---LAAGESPAMPDLVSKVTFPGSGGVAVSVSHKICIDIASLVTFVKDWAATTAKG 174
QY 181 EN-----IHPSLSEALLPFRD---DLPEKYVDQMEALWFAKVKVATRRFVFG 228
Db 175 SNSTIEAEITVPPPSHMYEQPSTSDSNITSYV-----LKRFPPEP 220
QY 229 KAISIODEAKESVPKPSRVHVTGFLWKHLIAASRALTSGTSTRLSIAAQVNLKTR 288
Db 221 SKIAELKHAASESVVPVTRVEAIMSLIWR---CARNSRSNLLIPRQAVMWAQMDCLR 277
QY 289 WMETVLDNATGNLFWNAQALLESHTTPEISDLKCDLVNLLNGSVKQCGNDYFETFKG 348
Db 278 IPSVVAQDVIGNL-----QSGFSLKKDAB--SEFEIETVATFRKNERNEMKESLQ 331
QY 349 KEGYGRMCEYLDLDFORTSMSEPPADYILFSSWTFN-FNPLDFGWRGTSMIGVAGKIESAS 407
Db 332 NT-IGQSLLSMAETVSESTE-IDRYIMSMCKPKPFVEVDFGSGSPVWVGYSHTIYDN 388
QY 408 CKFIILVPTQCGSGIEAWNVNLEEKAMLEQDPHFLALA 446
Db 389 MVGVVLIDSKEGDGEAVMSLPEEDMSVFVDDQELLAYA 427
RESULT 4
T46216
hypothetical protein T8P19.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 17-Mar-2000
C:Accession: T46216
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa
submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23008

A:Accession: T46216

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-430 <CHO>

A:Cross-references: EMBL:AL133315

A:Experimental source: cultivar Columbia; BAC clone T0P19

C:Genetics:

A:Map position: 3

A:Introns: 144/3

A:Note: T0P19.230

C:Superfamily: Arabidopsis CER2 protein

Query Match 11.0%; Score 262; DB 2; Length 430;

Best Local Similarity 24.5%; Pred. No. 5.7e-14;

Matches 115; Conservative 72; Mismatches 176; Indels 106; Gaps 23;

```

QY 16 PPSSTPLQYKLTLDQTPPAYVPIVFPY-----ITDHPNLQTLADLQALSETLT 71
DB 18 PPVSETPNGHYLSNLDQ-NIAIIVKTLVYKSGESRTNQESYNV-----IKKSLSEVLV 70
QY 72 LYVPLSGRV-----KNMLYIDDFEGVPLEARVNC---DMTDLRLKIECIENEFYPIKP 124
DB 71 HYPVAGRLTISPEGKIANVCTGEGVVVVEAEANCIGDITIKKAISENRMETLEKLYVDVP 130
QY 125 FSEMAISDERYPPLGVQVNVFD-SGIAIGVSVSHKLDGTCADCFKLSWGAVERGCRENI 183
DB 131 GARNIL---EIPPVVQVNTFKCGFVILGLMSHNFEDGVAABEFLNSWCMAKGLPLSV 187
QY 184 IHPSLSEALL---FPPRDLPEKYVDQMALWPAK-----KVATRRFVFGVKAISIQ 235
DB 188 --PPFLDRLTILGRNPKIEFPNHPEDIEDISTGKIYDEEKLIVKSLFEPFEKLEK 245
QY 236 DEAKSE-SVPKPSRVHATGFLKGLIAASEALTGTSTTRLSIAAQAQVNLRTM----- 289
DB 246 IMAIEENNNKYSTFOALTGFLKRSCEALR--FKPDQRYKLLFAADG---RSRFIPRLP 300
QY 290 -----NMETVLNDNATGLFWAAQAILLSHTTPEISDLKCDLVNLLNGSVKQC 338
DB 301 QGYCGNGIVLTGLVTSSEGLVGN-----PLSHSVGLVK-----RLVELVTDGFWRS 346
QY 339 NGDYFTFKGEGYGRMCEYLDQRTMSMEPAPDIYLFSSWTNF-FNPLDFQWGRITSW- 396
DB 347 AMDYFEV-----NRTSPSNAT---LLITSWSKLTLLKLDQFGWGEVPS 387
QY 397 --IGVAGKIESACKFILLVPTQCGSG-----TEAWNLSEKXAMLEQ 438
DB 388 GPVGLPGR-----EVLFLP-----SGDDMKSIINVFLGPTSAMVFEER 426

```

RESULT 5

T10711

anthranilate N-benzoyltransferase (EC 2.3.1.144) - clove pink

N:Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase

C:Species: Dianthus caryophyllus (clove pink)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T10711

R:Yang, Q.; Watern, U.; Grimmig, B.

Plant Mol. Biol. 38, 1201-1214, 1998

A:Title: Anthranilate N-hydroxycinnamoyl/benzoyltransferase gene from carnation: Rapid e

A:Reference number: Z17092; MUID:99084770; PMID:9869425

A:Accession: T10711

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-446 <YAN>

A:Cross-references: EMBL:Z98758; NID:93288179; PIDN:CAB11466.1; PID:g3288180

C:Function:

A:Description: catalyzes the synthesis of anthranilate

A:Pathway: phytoalexin biosynthesis

C:Keywords: acyltransferase; coenzyme A

Query Match

Best Local Similarity 23.3%; Score 242; DB 2; Length 446;

Matches 114; Conservative 88; Mismatches 185; Indels 102; Gaps 26;

```

QY 4 IEVSINSKHTIKSTSTPLQYKLTLDQ--TPPAYVPIVFPYPTDH----- 51
DB 1 MSIQIKOSTWREFA-BETPNKSLWLSKIDMLIRPYSHTGAVLIYKQPDNNEDIHSSSS 59
QY 52 ---DFNLQTLADLQALSETLTLYYPLSGRVKNN---LYIDDFEGVPLEARVNCMT 105
DB 60 MYFDANI-----LIEALSALVPIYYPWAGRLKINGDRYEIDCNAGALFVEAESHVLE 113
QY 106 DFLRLARKIECINE-FVPIKPFSEMAISDERYPPLGVQVNVFD-SGIAIGVSVSHKLDIGG 163
DB 114 DFGDFRPNDLHRVMVFTCDYS-KGISS--FPLIMVQLTRFCGVSIGFQAQHEHACDGM 170
QY 164 TADCFKLSWGAVERGCRENI--IHPSLSEALLFEPR-----DOLPEKYVDQM 209
DB 171 SHEEFNNWARIKAGLLPALPEVHRYLHLRLNPPQIKVTHSQEPFVPSLPNELLD-- 228
QY 210 EALWFAAGKIVATRR-FVFGVKAISIQD--EAKSESVKPSRVHATGFLWKHLIAASRA 266
DB 229 -----GKTNKSQTLFKLSREQINTLQKLDLSSNTTTLSTYEVVAGHVWRS-VSKARG 281
QY 267 LTSGTSTRLSIAAQAQVNLRTMNMETVLNDNATGLFWAAQAILLSHTTPEISDLKCD 326
DB 282 L-SDHEEIKL---IMPVDGRSRINNPSPKGYCGN-----VFLAVCTATVGDLS-CN 329
QY 327 LVNLLNGSVKQCGDYFETFKGEGYGRMCEYLDQRTMSMEPAPDI----- 374
DB 330 PLTDTAGKVQ-----EALKGLDD-----DYL--RSAIDHTESKPLPFPYMSPEKTL 375
QY 375 ---YLFSSWTNF-FNPLDFQWGRITSWIGVAGKIESACKFILLVPTQCGSG-IEAWNL 429
DB 376 YPNVLANSWGRIPVQAMDFGWSPTFGISNIFYDGC---FLIPSQNGDSMTLAINLF 432
QY 430 EEXXAMLEQ 438
DB 433 SSHLSLFFK 441

```

RESULT 6

T00918

hypothetical protein F21B7.32 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999

C:Accession: T00918

R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Co

eologis, A.; Ecker, J.R.

submitted to the EMBL Data Library, January 1998

A:Description: Genomic sequence for Arabidopsis thaliana BAC F21B7.

A:Reference number: Z14208

A:Accession: T00918

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-461 <SHI>

A:Cross-references: EMBL:AC002560; NID:92618677; PID:g2809263; GSPDB:GN00059; ATSP:F21B

C:Genetics:

A:Gene: ATSP:F21B7.32

A:Map position: 1

Query Match

Best Local Similarity 10.2%; Score 242; DB 2; Length 461;

Matches 120; Conservative 76; Mismatches 169; Indels 108; Gaps 27;

```

QY 4 IEVSINSKHTIKSTSTPLQ-----YKLTLDQTLTPPAYVPIVFPYPTDHFNLQ 57
DB 15 IPVTINQOVLVHFS-SPTPANQSPHSLYSLNLDODIGARVFTPSVFPYSTNREST-- 71
QY 58 TLADLQALSETLTLYYPLSGRVK--NNLYIDDF---EEGVPLEARVNCMTDQFLRLK 112
DB 72 VLKELQALSEVLVYPYPLSGRLREVNGKLEVEFFGEQGVLMVNSANSSMDLAD----- 125
QY 113 IECLNEFVPIKPFSEMAI-----SDERY-----PILGVQVNVFD-SGIAIGVSVSHKLDIG 162

```

Db 126 ---LGLDTPVNPAMPLIFRNPGBEAYKILEMPLLIAQVTFPTCGFSLGRLCHCICDG 182
Qy 163 GTADCLKSGAVRGCRGRENIIHPS-----LSEALLFPDRDLP-EKYVDQWEA 211
Db 183 FGAMQFGSGWAATAK-TGKLIADPEPWDRFTKPRPPPMVKYPHHEVLPPIERSNLTNS 241
Qy 212 LWFAGKKV-----ATRRFVFGVKAISSTQDEAKSESVPKPSRV-----HAVTGFLEKHLIA 262
Db 242 LMDT-KPLQKCYRISKFCVCKSIAGQED-----PTLVCSSTDMAAAHWRSWV- 290
Qy 263 ASRALTSSTSTRLSIAAQAVALNTRMNETVLDNATGNLFWAAQAIL-----ELSH 314
Db 291 --KALDVKPLDYNLRLETF-SYVNRLETLKLRKGFYGVVCLACAMSSVESLINDLSLK 347
Qy 315 TTPRISDLKCLDLVNLNGSVKQCGNGDYFETFKSGEGVGMCEYLDQRTWSSMEPAPDI 374
Db 348 TTRLVQDARL-----RVSEYILRS-----MVDIVDWKRP-KRLEFGK 385
Qy 375 YLFSSWTNP--FNPLDFGWRGTSWIGVAGKIE-SASCKFIILVPTQCGSGIEA 424
Db 386 TI-TOWTRFEMETADFGWCKPVY--AGPIDLRPTQVCVLLPQ---GGVES 431
RESULT 7
T51458
hypothetical protein K10A8_20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51458
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <SAT>
A;Cross-references: EMBL:AL391151
A;Experimental source: cultivar Columbia; BAC clone K10A8
C;Genetics:
A;Map position: 5
A;Introns: 147/3
A;Note: K10A8_20

Query Match 9.9%; Score 234.5; DB 2; Length 461;
Best Local Similarity 24.7%; Pred. No. 1.3e-11;
Matches 106; Conservative 69; Mismatches 178; Indels 77; Gaps 20;
Qy 39 YVPIVPEY---PITDHPNLPOTLADIQALSETLTYLYPLSGRVK-----NNLYIDFEE 91
Db 42 HIPTIFPYRNPTNSD-----PVAVIRRALAETLVYYPFAGRLREGPNRKLAVDCTGE 96
Qy 92 GVPYLEARVNCMTDFLRLKIE---CLNEFVPIKPFMSMAISDE-RYPLIGVQVNVFD 146
Db 97 GVLFIADADVLTVSEFEKDALKPPFCFPELL-----FNVGSCMLNTPMLMQVTRLK 152
Qy 147 -SGIAIGVSVSHKLIIDGTADCLFKSGAVRGCRGRENIIHP-----SLSEALLFPDRD 200
Db 153 CGGFIFAVRINHAMSAGGLTLFLKTCFVRGYHATVPVWERHLSARVLL---RVT 209
Qy 201 LPEKYVDQWEALFAGKVVATR-----FVGVKAISSTQDEAKSESVPKPSRVNAV 253
Db 210 HAHEYDEMPAI---GTEIGSRDNLVGRSLFFGFCMSAIRLLPPLNLSNSTNMEMLT 266
Qy 254 GFLWKHLIAASRALTSSTSTRLSIAAQAVALNTRMNETVLDNATGNLFWAAQAILLS 313
Db 267 SPLWRYETIALR--PDQKEXRLIL---IVNRSKLKNPPLPRGYGNAPFVVAIA--- 318
Qy 314 HTTPEISDLKCLDLVNLNGSVKQCGNGDYFET-----FKGEGYGRMCEYLDQRTWSS 367
Db 319 -TANELTKPLESALRLIKEAKSVTEYMRSLADIAMVIGRPSFSDGAYL----- 369
Qy 368 MEPAPDIYLFSSWTFNFPNPLDFG-WGRTSW--IVGAKIESASCKFIILVPTQCGS-GIE 423

Db 370 ---VSDVRIFAD-----IDFGIWGKPVYGGIGTAGVEDLFGASFYVSEKRNGEIGIV 419
Qy 424 AAWNLEEEKM 433
Db 420 VPVCLPEKAM 429
RESULT 8
T10719
anthranilate N-benzoyltransferase (EC 2.3.1.144) (clone pchcbr3) - clove pink
N;Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C;Species: Dianthus caryophyllus (clove pink)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10719
R;Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.
Plant Mol. Biol. 35, 777-789, 1997
A;Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA: a
lus L.
A;Reference number: Z17095; MUID:98088004; PMID:9426598
A;Accession: T10719
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-445 <YAN>
A;Cross-references: EMBL:Z84571; NID:g2239090; PIDN:CAB06538.1; PID:g2239091
C;Function:
A;Description: catalyzes the synthesis of anthranilate
A;Pathway: phytoalexin biosynthesis
C;Keywords: acyltransferase; coenzyme A

Query Match 9.8%; Score 233.5; DB 2; Length 445;
Best Local Similarity 23.5%; Pred. No. 1.5e-11;
Matches 114; Conservative 92; Mismatches 189; Indels 91; Gaps 27;
Qy 4 LEVINSKHTIKSTSTPLQPKYKLTLLDQL--TPPAVVPVIVFVPIITDHPNLPCLILAD 61
Db 1 MSIHIKOSTWVRPA-EETPNKSLMISKIDMLIRTPSYHTGAVLTKQPDNNENIQSSS 59
Qy 62 -----LQALSETLTYLYPLSGRVKNN---LYIDDPEEGVPLEARVNCMTDFLRLR 111
Db 60 MYFDANILIEALSALVPIYPMAGRLKINDRYEIDCNGEGALFVEAESSHVLEDFGDFR 119
Qy 112 KIECLNE-FVPIKPFMSMAISDERYPILLGVQVNVFD-SGIAIGVSVSHKLIIDGTADCL 169
Db 120 PNDELHRVMVPTCDYS-KGISS--FPLIMVQLTRFCGGVSGIGFAQHVVCDRMSHFEEN 176
Qy 170 KSMGAVPRG-----CRENIH-----PSLSEALLFPD-RDDLPEKYVDQWEALWFA 215
Db 177 NSWARIANGLLPALEFVHDHYLHLCPRNPQIKYTHSQFEPFVPSLFEKLLD----- 228
Qy 216 GKQVATRR-FVFGVKAISSTQDEAK-SESVKPSRVHVAVTGFLWKHLIAASRALTSSTTS 273
Db 229 GKTSKQTLFKLSSEQINTLKQKLDWNTTTTLSTYEVVAGHVWS-VSKARGL-SDHEE 286
Qy 274 TRLSIAAQAVALNTRMNETVLDNATGNLFWAAQAILSETTEISDLKCLDLVNLNG 333
Db 287 IKL---IMPVDEGSRINNPSPKPGYCGN-----VVFLAVCTATVGDLE-ACNPLETDTAG 335
Qy 334 SVKQCGNGDYFETFKGEGYGRMCEYLDQRTWSSMEPAPDI-----YLF 378
Db 336 KVQ-----EALKGLDD-----DYL--RSAIDHTSKPDLIPVPMGSPKLTLYPNVLN 381
Qy 379 SWTNF-FNPLDFGWRGTSW-GVAGKIESASCKFIILVPTQCGSG-IEAWNLEEKRAML 436
Db 382 SWGRIPYQAMDFGNGNPTFFGISNIFYDGC---FLIPSQNGDGSMTLAINLPSSLSLF 438
Qy 437 EQDPHF 442
Db 439 KK--HF 442

RESULT 9
T10718
anthranilate N-benzoyltransferase (EC 2.3.1.144) (clone pchcbr3) - clove pink (fragment)

N;Alternate names: anthranilate N-hydroxycinnamoyl/benzoyltransferase
C;Species: *Dianthus caryophyllus* (clove pink)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10718
R;Yang, Q.; Reinhard, K.; Schiltz, E.; Matern, U.
Plant Mol. Biol. 35, 777-789, 1997
A;Title: Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA:anthranilate L.
A;Reference number: Z17095; MUID:98088004; PMID:9426598
A;Accession: T10718
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-442 <YAN>
A;Cross-references: EMBL:Z84384; NID:G2239084; PIDN:CAB06428.1; PID:G2239085
C;Function:
A;Description: catalyzes the synthesis of anthranilate
A;Pathway: phytoalexin biosynthesis
C;Keywords: acyltransferase; coenzyme A

Query Match 9.8%; Score 232; DB 2; Length 442;
Best Local Similarity 23.6%; Pred. No. 1.9e-11;
Matches 109; Conservative 82; Mismatches 173; Indels 98; Gaps 25;

QY 14 IKPST-----SSTPLQPKYLTLDLQ--TPAYVPVIVFFYPITDH-----DF 53
DB 2 IKQSTWVRPAEETPKSLWLSKIDMILRTPYSHTGAVLIYQPDNEDNIHPSSMYFDA 61
QY 54 NLPQTLDLQALSETLTYPLSGRVKN---LYIDDFEGVPYLEARNVCDMTDFLR 110
DB 62 NI-----LIEALSALVPPVPMAGRLKINGDRYEIDCNAGALFVEAESHVLEDFGDF 115
QY 111 RKECLNE-FVPKPPSMEALDERYDILGVQNVFD-SGIAIGVSVSHKLDGTDACP 168
DB 116 RPNDLHRVMVPTCDYS-KGISS--FPLLMVQLTRFCGGVSGIQAQHHVCDGWAHFF 172
QY 169 LKSGWAFRG-----CRNLIH-----PSLSEAALLPPP-RDDLPEKYVDQMEALWF 214
DB 173 NNSWARIAGLLPALEPHRYHLRPNPPIQKSHSQEPFVPSLNNLLD----- 225
QY 215 AGKQVATRR-FVGVKAISIQDEAK-SESVKPSRVHATVGLFWKHLIAASRALTSQT 272
DB 226 -GKTNKSQTLLSREQINTLKQKLDLSNNTRELSTVEVAGVWRS-VSKARGL-SDHE 282
QY 273 STLSIAAQAQVNLTRMMETVLNATGNLFWAAQALIELSHHTTPEISDLX---LCOLVN 329
DB 283 EIKL---IMPVDGSRINNPSPKSGCN-----VFLAVCTATVDGUSCNELTDTAG 332
QY 330 LLNGSVQKQNDVPETFKGEGYGRMCEYLDLDFQRTSMSSM-----EPAPDIY---LFSSW 380
DB 333 KVQELAKGLDDYLRS-----AIDHTESKPGLPVPYMGSPKTLVPPVNLNSW 380
QY 381 TNF-FNPLDRGWRGTSVIGVAGKTESASCKFIILVPTQCGSG 421
DB 381 GRIPYQAMDFGWSGPTFFGINSIFDQGC---FLIPSRDGDG 419

RESULT 10
T00527
hypoethetical protein At2g19070 [imported] - Arabidopsis thaliana
N;Alternate names: hypoethetical protein T20K24.8
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 02-Feb-2001
C;Accession: T00527; B84572
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, July 1997
A;Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A;Reference number: Z14167
A;Accession: T00527
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-451 <ROU>
A;Cross-references: EMBL:AC002392; NID:g3176701; PID:g3176709
A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L. euss, D.; Niemman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.D.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: B84572
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-451 <STO>
A;Cross-references: GB:AE002093; NID:g3176709; PIDN:AAD1025.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g19070; T20K24.8
A;Map position: 2
A;Introns: 322/1

Query Match 9.6%; Score 228.5; DB 2; Length 451;
Best Local Similarity 22.6%; Pred. No. 3.9e-11;
Matches 99; Conservative 70; Mismatches 182; Indels 87; Gaps 19;

QY 13 TIKPSTSTPLQ-----YKLTLDLQTPPAYWIVFFYPITDHFNLPQTLADLRQALS 67
DB 5 TERKSYTIVFAEPTWSGRFPPLAEWCQQTTHITHTLYFYDKPSSSQ-GNVVEILKTSLS 63
QY 68 ETLTYPLSGRV---KNLYIDDFEGVPYLEARNVCDMTDFLRKIECLNEFVPIK 123
DB 64 RVLVHFTPMAGRLKMLPRGRFELNCNAGVEFIAESEGKLSDF-----KDFSTPT 114
QY 124 PFS--MEAIS---DERYPLLGQVNVFD-SGIAIGVSVSHKLDGTDACFLKSGWAVF 176
DB 115 EFENLMPQVYKNFIETIPLFLAQVTKFGKGSLSVNVSHAVDQGSALHISEWGLA 174
QY 177 RGRENIIHPSLSAAL-----LPPPR-----DDLPEKYVDQMEALFAGKVKVATR 222
DB 175 RGEPLTV-PFLDRKILWAGEPLPPFPVSPFPDKFEDQPPFLIGETDNVERKKKTTV 233
QY 223 RFVGVKAISIQDEAKSESVPKP-----SRVHATVGLFWKHLIAASRALTSQTSTRLSI 278
DB 234 MLPJSTSQLKRSKANGSKHSDSAKGFTRVETVGVHVR---CACKARGHSPQPTALGI 291
QY 279 AAQAVNLRTNMETVLNATGNLFWAAQALIEL--SHHTPEISDLKLCDLVNLINGSVK 336
DB 292 C-----IDTRSEMPPLRG-----YFGNATLDVVAASTSGELISNELGFAASLIKAIK 341
QY 337 QCNGDYETFKGEGYGRMCEYLDLDFQRTSMSSMEAPDIYLFSSWNTFF--NP----- 386
DB 342 NVINETYMI-----GIYLNQKDLKRFQ---DLHALGSTGEPFYGNPNLGVWSWL 389
QY 387 -----LDFGWRGTSWIG 398
DB 390 TLENYGLDFGWSGKEFYTG 407

RESULT 11
G86453
YUP8H12R.39. homolog P9111.9 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G86453
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K. ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86453
A;Status: preliminary
A;Molecule type: DNA

```

Qy 52 ---DFNLPTQLADLQAALSETLTLYPLSGRVKNN---LYIDDFEGVGVYLARVNCIDMT 105
Db 60 MYPDANI-----LIEALSKALVPFPMAGRLKINDRYEIDCNAGCALVEAESHVLE 113
Qy 106 DFLRLKIECLNE-FVPIKPFSMEAISDBERYLLGVQNVFD-SGIAIGSVSVSHKLIDGG 163
Db 114 DFGDFRNDLHRVMVPTCDYS-KGISS--FPILMVOLTRFCRGVSGISGEAQHHVCDDGM 170
Qy 164 TADCFLSKGAVFRG-----CRENIH-----PSISEAALLFP-RDDLPKYVDQM 209
Db 171 AHFFNNSWASIAKGLLPALBPVDRYHLPRNPPOIKYSHSQCFEFVPSLPELLED-- 228
Qy 210 EALWFAGKKVATR--FVFVGKRAISSIQDEAK--SESVPKPSRVHAVTGFLWKHLIAASRAL 267
Db 229 -----GKTNSKQTLPILSRBOINTLKOKLDLSNNTTRLSTYEVAHVWRS-VSKARGL 281
Qy 268 TSGTTTSLSTAAGANVRTRNMNETVLDNAENLFWWAQAILLESHTTPEISDK--L 324
Db 282 -SDHEEIKL---IMPVDGRSRINPSPKGYCGN-----VVFLAVCTATVGDLSNPL 330
Qy 325 CDLVNLLSGVKQCNGDYFENFKGEGYCRMCEYLDLFQRTMSM-----EPAPDIY--- 375
Db 331 TDTACKVQBALKGLDDDDYLR-----AIDTESKPGCLPVPMYGSPEKTLIPNV 378
Qy 376 LFGSWTWTF-FNPLEPWGWRTSGIWGAKIESASCIFILLVPTOCGSG 421
Db 379 LVNSWGRIPTYQAMPFGMGSPTFPGFINPYDGC--FLIPSRDGDG 422

RESULT 13
T01056
hypothetical protein YUP8H12R.39 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 22-Oct-1999
C;Accession: T01056
R;Theologin, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; Kwan
Cefner, P.; Davis, R.W.
submitted to the EMBL Data Library, May 1998
A;Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.
A;Reference number: Z14227
A;Accession: T01056
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-572 <THE>
A;Cross-references: EMBL:AC002986; NID:g2494106; PID:g3152598; GSPDB:GN00059; ATSP:YUP8H
C;Genetics:
A;Gene: ATSP:YUP8H12R.39
A;Map position: 1
A;Introns: 71/2; 98/1; 257/3

Query Match          9.4%; Score 222; DB 2; Length 572;
Best Local Similarity 22.9%; Pred.No.1.9e-10;
Matches 115; Conservative 86; Mismatches 159; Indels 142; Gaps 28;

Qy      6 VSINSKHITKISTSTGPCLKVLT--LDQLTPPAVVPVFEVFIPTDHDNLPTLA--D 61
Db     130 VTTRKQVI---TAALPCDHWLPUSNLDLLPLPNVHVCFCKKPLHFN---TVAYET 183

Qy     62 LRQALSETLTLYPLSGRVKNN-----LYIDDFEGVGPYLEARVNCIDMTDFLRKIEC 115
Db    184 LKTAALAEIIVSYYPAGELVTNETGPEPELCNN--RGVDFVEAGADVRELNLNYDPDES 241

Qy    116 LNEFVPKIPFSMEAISDBERYLLGVQNVFDGS-TAIGSVSVSHKLIDGGTADCFLSKWSGA 174
Db   242 IAKLVPIK-----KHGVIAIQVTKLKGSIWVGCTFDRHVADAYGMNMPFLSWEAE 291

Qy    175 VFR-----GC-----RENIHPSLSEAAAL----LPPRRDUL-PKHYVDOMEAL 212
Db   292 ISRSDVPSICVPSFARSLLNPRPLVMDPSIOQIYMPVTISLPPQETTNEPL----- 344

Qy   213 WFAGKKVATRFRFVGKVAISSIQDEAKSESVPKPSRVHAVTGF-WKHLIAASRALTSGT 272
Db  345 -----LASRIYYTKANALOLOTASSKNCKRTKLESFSAFMK-LVAFAAKDPVPI 397

```

QY 273 ST-RLSIAQAQVNLRTNNMETVLDNATGMLF---WVAQAILE-----LSHTTPEISDLK 323
DB 398 KTSKLG---VVDGRRL-MEKENNYFGNVLSPFGQGRIDDLISKPLSNVTEEVH---450
QY 324 LCDLVLLNGSVKQCGNDYFETFKGKGYGKMGCSYLDLQRTMSSMEPADPDI-----374
DB 451 ---REFKSVT-----KEHFLNLDWVE-----TCRTTPAVSRYSYSGSD 487
QY 375 ---YLFSSWTNF-FNPLDFGWG-----RTSNIGVAGKIESASCKFIILVFPQCGSGI 422
DB 488 DGAFAVYSSGRSPFNQVDFGWSGSPVFGSYHFPWGSAG-----YVWEMP-----SSV 535
QY 423 E-----AWVNLSEKMAMLQD 439
DB 536 DDRDMVYLHLTKGQLRFIEE 557
RESULT 14
G84823
probable anthocyanin 5-aromatic acyltransferase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84823
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
euss, D.; Nierman, W.C.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84823
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-482 <STO>
A;Cross-references: GB:AE002093; MID:g2088651; PIDX:AAB95283.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g39980
A;Map position: 2
Query Match 9.2%; Score 219; DB 2; Length 482;
Best Local Similarity 22.5%; Pred. No. 2.7e-10;
Matches 122; Conservative 93; Mismatches 171; Indels 156; Gaps 26;
QY 1 MEKLEVSINSKHTIKPSTSTPLQPYKLTLLDQLTPPAYVPLVFPYI-----TDHFN 54
DB 4 LEX-SVTLISGRNVFPQKST-LVDLKLVSVD-----LWLSCHYIQKCLFTCPNLP 54
QY 55 LPQTLADLRQALSETLTYPLSGRV-----KNLYIDDPEEGVPYLEAR-----V 100
DB 55 LPALISHLKHLSLTILTHPPLAGRLSTSSSGHVFELCNDAQDFVFAQAKSIHVSDVIA 114
QY 101 NCDMTDFLRKIECLNEFVPIKFSME-ALSDRY-PLLGQVNVFDSGIAIGVSVSH 157
DB 115 GIDVPDVK-----EF-----FTYDRAVSVEGHNRPILAVQVTELDGCVFICGSVNH 161
QY 158 KLIDGGTADCFKSGAVFRCNENIHPSL-EEALLFPFRDDLPEKYVDQMEALWAG 216
DB 162 AVTDGTLNWFINTFAEVSRAKQWTPQDPTRESVLISPAVLKVPQ-----GG 210
QY 217 KKVA-----TRFVFGVKAISIQ-----DE-----237
DB 211 PKVTFDENAPLRLRIFSRRESIQELKAVNKKWLTVDNGEIDGVLLGKQNDKNGK 270
QY 238 -----AKSESVKPK-----SRHAVTGFLWKHLIAASRALTSCTTSTRSIA 279
DB 271 ENGILTEMLSELFGRNDVAVSPVAVE-SSFSQSCALLWRAITRAKLPSSKTTTFRM---327
QY 280 AQAVNLRTNME-----TVLDNATGNLFWAAQALIELSHTTPEISDLKLC-DLVNL-----330
DB 328 --AVNCRHRLSPKLNPEYFGNAIOSVPTFATAAEVLSR-----DLKWCADQLNQSVAAH 379
QY 331 LNSGVKQCGNDYFETFKGKGGMCEYLDLQRTMSSMEPADPDIYLFSSWTNFNPLDFG 390

DB 380 QDGRIRSVADW-----EANPRCFPLGNADGASVTMGSSPRFPMYDN-----DFG 424
QY 391 WGRTSWI--GVAGKIESASCKFIILVFPQCGSG-IEAVNLEEKMALEODPHFLALAS 447
DB 425 WGRPVAVRSGRSNKFDGKISAF-----FCREGNGTVLDLEWILSPETMAGIESDGBFMYVT 480
QY 448 PK 449
DB 481 NK 482
RESULT 15
T45611
N-hydroxycinnamoyl/benzoyltransferase-like protein - Arabidopsis thaliana
N;Alternate names: protein F13G24.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45611
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voest, M.; Robben, J.; Volckaert, G.; B
submitted to the Protein Sequence Database, December 1999
A;Reference number: 223009
A;Accession: T45611
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-454 <BEV>
A;Cross-references: EMBL:AL133421
A;Experimental source: cultivar Columbia; BAC clone F13G24
C;Genetics:
A;Map position: 5
A;Introns: 159/3
A;Note: F13G24.60
Query Match 9.1%; Score 217; DB 2; Length 454;
Best Local Similarity 22.6%; Pred. No. 3.7e-10;
Matches 113; Conservative 82; Mismatches 191; Indels 114; Gaps 23;
QY 5 EVSINSKHTIKPSTSTP-LQPYKLTLLDQLTPPAYV---PIVFYP-----ITDHDENL 55
DB 7 EVKLISKCFVKPKTIPEKWKPEYHFSMDHVLISHYIQKGLLFLKPSFSESVTPKEF-M 65
QY 56 PQTADLRQALSETLTYPLSGRV-----KNLYIDDPEEGVPYLEARNVCMMD 106
DB 66 ETLQKLKSLAIALVHFYPLAGISITLKTNDSSRSVDFVDCNNSPAGFIAESDLSVD 125
QY 107 FLRLKIECLNEFVPIKFSM-----EALS--DERVPLAGVNVFDSGIAIGVSVSHKI 160
DB 126 ILG-----SKYVPLVWQSFDDHKKALSRGDTMTLLSVKVTLDVGVFGLSNHSLG 178
QY 161 DGGTADCFKSGAVFRCNENI-----IHPSEEAALLFPFRDLPPEKYVDQMEA 211
DB 179 DGSSFWHFFNSLSLEFNSQEDNNKFLCKNPPIFREVSGPMVSLPFE--PDESISQSER 236
QY 212 LWFAGKVATRRFVFGVKAISIQDEAKSE-SVPKPSRVHVTGFLWKHLIAASRALTS 270
DB 237 ---PVLKERMFHFSSETVRSLSKANEEOGTTKISSLSQSLTALIMRSITRAKLPNDQ 291
QY 271 TTSTELSLAAQAVNLRTNME-----TVLDNATGNL-----FWA-----QAI 309
DB 292 ETTCLRL-----AAGNRSRMNPLPNHFGNISLVIAITTTGDLLENFGCAALKHQAV 346
QY 310 LEHSHTTPEISDLKLCIDLVNLLNGSVKQCGNDYFETFKGKGGMCEYLDLQRTMSSME 369
DB 347 TE--HTGEKIS--ADMWRWLKAHLK-----LDGF-----FSPNIVHMG 380
QY 370 PADIYLFSSWTNFNPLDFGWRGRTSMI--GVAGKIESASCKFIILVFPQCGSGIEAVN 427
DB 381 SSPRENKYGS-----EFGMGKAVAVRSVGGKYGIDG---KVSAYPREGGASIDLEYC 429
QY 428 LEEKMALEODPHFLALAS 447
DB 430 LPPECMALELDQEFMSLVS 449

Search completed: July 6, 2004, 13:39:30
Job time : 21.1722 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:59 ; Search time 11.6195 Seconds
(without alignments)
2025.529 Million cell updates/sec

Title: US-09-857-518A-6
Perfect score: 2373
Sequence: 1 MEKIEVSINSKHTKESTSS.....MAMLEQDPHFLALSPKTLI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	177.5	7.5	441	1	DENT TAXCA
2	156	6.6	439	1	TSAT TAXCH
3	148.5	6.3	440	1	DBAT TAXCU
4	144	6.1	439	1	TSAT TAXCU
5	143	6.0	469	1	ANTA GENTR
6	142.5	6.0	440	1	DBBT TAXCU
7	102.5	4.3	900	1	SYA MYCPN
8	99.5	4.2	831	1	PLRLR MELGA
9	99.5	4.2	2607	1	BACB BACLI
10	99	4.2	574	1	IPA4 SHIEL
11	99	4.2	1281	1	MDR3 CRIGR
12	96.5	4.1	855	1	XAB2 HUMAN
13	96.5	4.1	855	1	XAB2 MOUSE
14	96.5	4.1	855	1	XAB2 RAT
15	95.5	4.0	2227	1	POLG HPAVH
16	95.5	4.0	2227	1	POLG HPAVH
17	95	4.0	505	1	Y4XG RHLSN
18	95	4.0	928	1	NRPI XENLA
19	94.5	4.0	2227	1	POLG HPAVL
20	93	3.9	6486	1	TYCC BREPA
21	92.5	3.9	741	1	LOXB PHAVU
22	92.5	3.9	2226	1	POLG HPAV2
23	92.5	3.9	2226	1	POLG HPAV4
24	92.5	3.9	2226	1	POLG HPAV8
25	91.5	3.9	397	1	V506 ROTPG
26	91.5	3.9	693	1	EX70 DROME
27	91	3.8	2230	1	POLG HPAV5
28	90.5	3.8	1056	1	POL BIV27
29	90	3.8	886	1	LEUR YEAST
30	90	3.8	1276	1	MDR2 MOUSE
31	90	3.8	1278	1	MDR2 RAT
32	88	3.7	709	1	KSP1 SCHPO
33	87.5	3.7	397	1	V506 ROTPY

34	87.5	3.7	404	1	GYRB BACMY	Q9x3y5 bacillus my
35	87.5	3.7	539	1	RN37 MOUSE	Q925f4 mus musculus
36	87	3.7	660	1	YJ42 CHROVO	Q7nwn9 chromobacte
37	87	3.7	1905	1	TAGB DICI	P54683 dictyosteli
38	86.5	3.6	730	1	EF2 MBTWT	Q93637 methanococc
39	86.5	3.6	1024	1	POPC RALSO	Q9rbs2 caulobactera
40	86	3.6	349	1	MOAA CAUCR	Q9ac48 streptococc
41	85.5	3.6	325	1	IADI STRP3	Q8k654 streptococc
42	85.5	3.6	397	1	V506 ROTHM	P03530 human rotav
43	85.5	3.6	404	1	GYRB BACCE	Q9x3v3 bacillus ce
44	85.5	3.6	522	1	PLIN HUMAN	O60240 homo sapien
45	85.5	3.6	578	1	SYD UREPA	Q9pdk5 ureaplasma

ALIGNMENTS

RESULT 1
DBNT TAXCA STANDARD; PRT; 441 AA.
AC Q8LL69;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase (EC 2.3.1.-)
DE DBTWTBT).
GN TAX10.
OS Taxus canadensis (Canadian yew).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Taxaceae; Taxus.
ON NCBI_TaxID=88032;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22103659; PubMed=12089320;
RA Walker K.; Long R.; Croteau R.B.;
RT "The final acylation step in taxol biosynthesis: cloning of the taxoid
RT C13-side-chain N-benzoyltransferase from Taxus.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9165-9171(2002).
CC -!- FUNCTION: Catalyzes the stereoselective coupling of the surrogate
CC substrate N-debenzoyl-(3'RS)-2'-deoxytaxol with benzoyl-CoA to
CC form predominantly one 3'-epimer of 2'-deoxytaxol.
CC -!- PATHWAY: Taxol biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the plant acetyltransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF466397; AA:75818.1;
DR InterPro; IPR003480; Transferase.
DR Pfam; PF02458; Transferase; 1.
KW Taxol biosynthesis; Transferase; Acyltransferase.
SQ SEQUENCE 441 AA; 49041 MW; 1F0C20CFDA6D4A0C CRC64;

Query March 7.5%; Score 177.5; DB 1; Length 441;
Best Local Similarity 24.3%; Pred. No. 2.3e-07;
Matches 108; Conservative 71; Mismatches 167; Indels 99; Gaps 24;
Qy 14 IKPSTSTPQPKYKLLDQLTDPAYVPIVFYPTDHDNLPTL-AD----LRQALSE 68
Db 19 VAPSLPS-PRATVQLSVDSLT----ICRGIFNTLL--VFNAPDNISADPVKIIREALSK 71
Qy 69 TLTLIYPLSGRVKN----NLVIDDFEGVPVLEARNVCMDFELRLKIECLN-----EFV 120
Db 72 VLIVYFFPLAGRLRSKBEIGEVECTGDALFVEAMVEDTIS---VLRLDLDLNFPSFOQLV 128
Qy 121 PIKFFSNEATSDERYPLIGVQVNVFD-SGIAIGVSVSHKLDGGTADCFKSWGAVFRG- 178
Db 129 FWHELD-TAIED--LHLIVIQVTRFTCGGIAVGVTLPVSHVCDGRGAAQFVTALAEARGE 185

QY	106	DFLRRLKIECLNFFVPIKP-----FSMEALSD-BRYPLLGVQUNVFD-SGIAIGVSUSH	157
Db	109	D-----NELSVLGDFDSDNPNFQQLFLSLPLDTNFKDLPLLVQVTRFTCGGVVGVSPFH	164
QY	158	KLIDGGTADCFKLSGWAVERGCRENIHPSLSSEALLFPDRDDLPEKVVVDQWEALWF---	214
Db	165	GVDCGRGAQFLKGLAMARG-----EVKLSLEPTINRELVLKDDPKYLOPHF	213
QY	215	-----AGKKVATRRFVFGVKAISSIQDEAKSSSVPKPSRVHVAVTGLWKHLIAASRAL	267
Db	214	EFLRAPSIVEKIVQTVFIIDFETINVIKQSVMECKEFCSSFEVASAMTW---IARTRAF	270
QY	268	TSGTTSTRLSIAQAQVNLRTFMNMTVLD-----NATGNLFWQAQILLSHTTPEISDL	322
Db	271	QIP-----RSEYVKILFGMDMRNENFPPLPSGGYGNSTGTACAV-----	309
QY	323	KLCDLVNLLNGS-----VQKCNQDVFETPKGKEGYGEMCEYLDQFQTMSSMEPA---	371
Db	310	--DNVQDLSSGLLRATMIILKSKVSLNDNPKSR-----AVVKESELDV	351
QY	372	-----PDYILFSSMTNF-FNPLDFQWGRTSWIG-VAGKTESASCK-FIILVPTQ-CGSGIE	423
Db	352	NMNHENVVAFADWSRLGDFEVDPCWGNVSVSPVQQCELAQNYVFLFKPSKPKPDGIG	411
QY	424	AWNLEEEKM	433
Db	412	ILMFLPLSKM	421

RESULT 3

ID	DBAT	TAXCU	STANDARD;	PRT;	440 AA.
AC	Q9W6R2;				
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	10-deacetylba	ccatin III 10-O-acetyltransferase (EC 2.3.1.167) (DBAT).			
OS	Taxus cuspidata	(Japanese yew).			
OC	Eukaryota;	Viridiplanetae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta;	Coniferopsida; Coniferales; Taxaceae; Taxus.			
OC	NCBI_TaxId=99806;				
RN	[1]				
RP	SEQUENCE FROM N.A. AND CHARACTERIZATION.				
RX	MEDLINE=20105515; PubMed=10639122;				
RA	Walker K., Croteau R.B.;				
RT	"Molecular cloning of a 10-deacetylba	ccatin III-10-O-acetyl			
RT	transferase cDNA from Taxus and functional	expression in Escherichia			
RT	coli."				
RL	Proc. Natl. Acad. Sci. U.S.A. 97:583-587 (2000).				
CC	-I- CATALYTIC ACTIVITY: Acetyl-CoA + 10-deacetylba	ccatin III = CoA +			
CC	ba	ccatin III.			
CC	-I- PATHWAY: Taxol biosynthesis.				
CC	-I- SUBUNIT: Monomer.				
CC	-I- SIMILARITY: Belongs to the plant acetyltransferase family.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce ,				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL; AF193765; AAF27621.1; -				
DR	PIR; T52320; T52320.				
DR	InterPro; IPR0031480; Transferase.				
DR	Pfam; PF02458; Transferase; 1.				
KN	Taxol biosynthesis; Transferase; Acyltransferase.				
SQ	SEQUENCE 440 AA; 49052 MW; 85CABE5010A68AC3 CRC64;				

Query Match 6.3%; Score 148.5; DB 1; Length 440;
 Best Local Similarity 23.6%; Pred. No. 6.8e-05;

ID SYA MYCPN STANDARD; PRT; 900 AA.
AC P75368;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase [EC 6.1.1.7] (Alanine--trna ligase) (Alars).
GN ALAS OR MPN419 OR MP422.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert R., Plagens H., Pirkel B., Li B.-C.,
RT Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
diphosphate + L-alanyl-tRNA(Ala).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; A300041; AAB96070.1; -.
CC PIR; S73748; S73748.
CC HAMAP; MF 00036; - 1.
DR InterPro; IPR002318; tRNA-synt_2c.
DR InterPro; IPR006193; tRNA_synt_ala.
DR Pfam; PF01411; tRNA-synt_2c; 1.
DR PRINTS; PR00980; TRNASYNTHALA.
DR TIGRFAMs; TIGR00344; alas; 1.
DR PROSITE; PS00860; AA tRNA LIGASE II ALA; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 900 AA; 103634 MW; 59C1A9FA0140B6E CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 900;
Best Local Similarity 23.6%; Pred. No. 1.6;
Matches 49; Conservative 32; Mismatches 80; Indels 47; Gaps 7;

QY 3 KIEVINSKHTIKPSTSTELQPKLILDLTPPAYVPIVFYPIITDHDENLPQTLADL 62
DB 719 KLSSTINSLSHTLNIANPALNKOKTALTQQL-----NHFHLQVITDL 762

QY 63 R-QALSETLLYYPLSGRVKNLYIDTFEGVPLEARVNCMDTFLRLRKIECLAEFF 120
DB 763 RKCQALLNELXI-----TVNELKTEDPKKQQLAEKIKQELLEAKQDKAYVLASFA 815

QY 121 PIKPFM-----EASIDREYPLGGVNVVD---SGIATGVSVS-----HKLDGGTA 165
DB 816 AVDPFKLLSQVAQAVLNQHNKFLVLLNQNPNNSPMLLQGVSKVCIQLKAKHFLKGGGS 875

QY 166 DCFKSKWGAVERGCRNIIHPSLSAAL 193
DB 876 NNF-----FRGSPNESVDVSKLQAIL 896

RESULT 8
PRLR_MELGA STANDARD; PRT; 831 AA.
AC Q91094; Q91091; Q91092;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Prolactin receptor precursor (PRL-R) (TPRLR).
GN PRLR.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97057891; PubMed=8902221;
RA Zhou J.F., Zadworny D., Guemene D., Kuhnlein U.;
RT "Molecular cloning, tissue distribution, and expression of the
prolactin receptor during various reproductive states in Meleagris
gallopavo";
RL Biol. Reprod. 55:1081-1090 (1996).
RN [2]
RP SEQUENCE OF 82-121 AND 473-522 FROM N.A.
RC TISSUE=Ovary;
RA Pitts G.R., You S.K., Foster D.N., el Halawani M.E.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This is a receptor for the anterior pituitary hormone
prolactin.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Belongs to the type I cytokine family of receptors.
CC Subfamily 1.
CC -!- SIMILARITY: Contains 4 fibronectin type III domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; L76587; AAB01544.1; -.
CC EMBL; U22947; AAA75038.1; -.
CC EMBL; U22924; AAA75039.1; -.
CC HSP; P16471; IEP3.
DR InterPro; IPR002996; CR1A.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003528; Hemtopopn_L_F1.
DR Pfam; PF00041; fn3; 4.
DR SMART; SMO0060; FN3; 3.
DR PROSITE; PS01352; HEMATOPO REC_L_F1; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
FT SIGNAL 1 23
FT CHAIN 24 831
FT DOMAIN 24 438
FT TRANSMEM 439 459
FT DOMAIN 460 831
FT DOMAIN 25 122
FT DOMAIN 123 225
FT DOMAIN 228 325
FT DOMAIN 326 428
FT DISULFID 36 46
FT DISULFID 75 86
FT CARBOHYD 59 59
FT CARBOHYD 91 91
FT CARBOHYD 100 100
FT CARBOHYD 112 112
FT CARBOHYD 132 132
FT CARBOHYD 262 262
FT CARBOHYD 303 303
FT CARBOHYD 315 315
FT CARBOHYD 335 335
SQ SEQUENCE 831 AA; 94394 MW; 2209163208777FAC1 CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 831;
Best Local Similarity 20.7%; Pred. No. 2.5; Indels 169; Gaps 21;
Matches 100; Conservative 57; Mismatches 169; Gaps 21;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 19-OCT-2003 (Rel. 42, Last annotation update)
 DE XPA-binding protein 2.
 GN XAB2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Kidney, and Stomach;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarrelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
 RA Villalon D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Involved in transcription-coupled repair (TCR) and
 CC transcription (by similarity).
 CC -1- SUBUNIT: Associates with RNA polymerase II, the TCR-specific
 CC proteins CKN1/CSA and ERCC6/CSB, and XPA (by similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Contains 11 HAT repeats.
 CC -1- SIMILARITY: Contains 10 TPR repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AK002890; BAB22435.1; -

DR EMBL; AK008628; BAB25790.1; -
 DR EMBL; BC021341; AAB21341.1; -
 DR MGD; MGI.1914689; 0610041014Rik.
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0005515; P:protein binding; ISS.
 DR GO; GO:0006283; P:transcription-coupled nucleotide-excision r. . .; ISS.
 DR GO; GO:0006350; P:transcription; ISS.
 DR InterPro; IPR003107; HAT.
 DR InterPro; IPR006940; Prenyl trans.
 DR InterPro; IPR008941; TPR-like.
 DR InterPro; IPR001440; TPR.
 DR SMART; SM00386; HAT; 10.
 KW DNA repair; Transcription; Nuclear protein;
 KW Repeat; TPR repeat.
 FT REPEAT 15 47 HAT 1.
 FT REPEAT 33 67 TPR 1.
 FT REPEAT 48 80 HAT 2.
 FT REPEAT 90 122 HAT 3.
 FT REPEAT 110 143 TPR 2.
 FT REPEAT 124 158 HAT 4.
 FT REPEAT 146 179 HAT 5.
 FT REPEAT 160 182 HAT 6.
 FT REPEAT 256 289 HAT 7.
 FT REPEAT 355 388 HAT 8.
 FT REPEAT 369 407 HAT 9.
 FT REPEAT 395 428 TPR 6.
 FT REPEAT 433 466 TPR 7.
 FT REPEAT 484 517 HAT 10.
 FT REPEAT 498 530 HAT 11.
 FT REPEAT 519 551 TPR 9.
 FT REPEAT 532 566 HAT 9.
 FT REPEAT 554 590 TPR 10.
 FT REPEAT 571 605 HAT 10.
 FT REPEAT 679 713 HAT 11.
 FT CONFLICT 684 684 A -> T (IN REF. 2).
 FT CONFLICT 842 842 Q -> L (IN REF. 1; BAB25790).
 SQ SEQUENCE 855 AA; 93987 MW; 6A5DA6A74E7FCH1D CRC64;
 Query Match 4.1%; Score 96.5; DB 1; Length 855;
 Best Local Similarity 21.1%; Pred. NO. 4.7;
 Matches 70; Conservative 49; Mismatches 114; Indels 99; Gaps 16;
 QY 167 CFLKSGAVFGCGRENILHPSLS-----EALLPFRDPLPEKYVDQMALWFA 215
 Db 66 CSYKLVRYLKARAQVGRVCTDPAYEDVNNCHERAFVFMHK--MPRLWLDYCOFLMDQ 123
 QY 216 GKQVATRR-FVFGVKAISSIQDE-----AKSESVKPSRVHVAVTGFLWKHLIAASR 265
 Db 124 GRVTHTRTFDRALRALPITQHSRIWPLYLRLRSHPLPE-TAVRGYRRFLKLPESAE 182
 QY 266 ALTSGLTSTRLSIAAQAVNLRTNMETVLDNACNLFWQAQAILLSHTTPTSDKLC 325
 Db 183 YIEYKASDRLDEAAQ--RLATVNDERFVSKA-----GKSNYQLWH-----ELC 225
 QY 326 DLVNLGSKVQKCGDYFEFTFKG-----KXGYGRM-CEYLDLQRTMSSMEPADIVY--- 375
 Db 226 DLISQNDKVKQSLNVD--AIIRGLTFTDLQSLKWLADYYIRSGHFEKARVISEAI 283
 QY 376 -----LPSSWTNFFNPLDFGNGRTSWIGVAGKIESASCKFIILVPTQCGSGIEA 424
 Db 284 RTVMTVRDFTQVDFSYAQFESM-----IAAKMETAS-----ELGREED 323
 QY 425 WVNLE-----EEKMA-----MLEQDPH 441
 Db 324 DVDELRLARFEQLISRRP-LILNSVLLRNPH 355
 RESULT 14
 XAB2 RAT
 ID XAB2 RAT STANDARD; PRT; 855 AA.
 AC Q99PK0;
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE XPA-binding protein 2 (Adaptor protein ATH-55).
 GN XAB2 OR ATH55.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RA Tezuka T., Sugai T., Amada N., Araki K., Nawa H.;
 RT "Cloning and characterization of a novel adapter protein ATH-55 in the
 PL brain".
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Involved in transcription-coupled repair (TCR) and
 CC transcription (By similarity).
 CC -!- SUBUNIT: Associates with RNA polymerase II, the TCR-specific
 CC proteins KKN1/CSA and ERC6/CSB, and XPA (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 11 HAT repeats.
 CC -!- SIMILARITY: Contains 10 TPR repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF277899; AAC53885.1; -;
 CC GO; GO:0005634; C:nucleus; ISS.
 CC GO; GO:0005515; F:protein binding; ISS.
 CC GO; GO:0006283; P:transcription-coupled nucleotide-excision r. . .; ISS.
 CC GO; GO:0006350; P:transcription; ISS.
 CC InterPro; IPR003107; HAT.
 CC InterPro; IPR008940; Preyyl trans.
 CC InterPro; IPR008941; TPR-like.
 CC DR SMART; SM00386; HAT; 10.
 CC DR Repeat; TPR repeat.
 KW DNA repair; Transcription; Nuclear protein;
 FT REPEAT 15 47 HAT 1.
 FT REPEAT 33 67 TPR 1.
 FT REPEAT 48 80 HAT 2.
 FT REPEAT 90 122 HAT 3.
 FT REPEAT 110 143 TPR 2.
 FT REPEAT 124 158 HAT 4.
 FT REPEAT 146 179 TPR 3.
 FT REPEAT 160 192 HAT 5.
 FT REPEAT 256 289 TPR 4.
 FT REPEAT 270 305 HAT 6.
 FT REPEAT 355 388 TPR 5.
 FT REPEAT 369 407 HAT 7.
 FT REPEAT 395 428 TPR 6.
 FT REPEAT 433 466 TPR 7.
 FT REPEAT 484 517 TPR 8.
 FT REPEAT 498 530 HAT 8.
 FT REPEAT 519 551 TPR 9.
 FT REPEAT 532 566 HAT 9.
 FT REPEAT 554 590 TPR 10.
 FT REPEAT 571 605 HAT 10.
 FT REPEAT 679 713 HAT 11.
 SQ SEQUENCE 855 AA; 99997 MW; 35D18C33F66121A9 CRC64;
 Query Match 4.1%; Score 96.5; DB 1; Length 855;
 Best Local Similarity 21.1%; Pred. No. 4.7;
 Matches 70; Conservative 49; Mismatches 114; Indels 99; Gaps 16;
 QY 167 CFLKSGAVFRGRENIIHPSLS-----EALLFPFRDLPPEKYVDQMEALWEA 215
 DB 66 CSYKLWRYLKARRAQVRCVTDPAVEDVNNCHERAFVMEK--MPRLWLDYQFLMDQ 123

QY 216 GKKVATPR_FVGVKAISSIOBE-----AKSESVKPKSRVHATVGLWKLHIAASR 265
 DB 124 GWTTRTFDRALRALPITQHSRWPLYLRLRSHPLPE-TAVRGYRRFLKLSPSAAE 182
 QY 266 ALTSCTTSLSLTAAQAVNLRTMMNMTVLDKATGNLFWAQAILELSTHTTPEISDKLC 325
 DB 183 YTEYLKSSDRLEDAQ--RLATVVDNERFVSKA-----GKSNYOLMWH-----ELC 225
 QY 326 DLVNLINGSVKOCNGDYFETPKG-----KEGYGRM-CEYLDFTQRTSMSEPAPDIY--- 375
 DB 226 DLISQPKDKVQSLNVD--ALIRGGLTFTDQGLKWLCSLADYVIRSGHFEKARDVVEAI 283
 QY 376 -----LFSWNTNFFNPLDFGNGRTSMIGVAGKIESASCKFIILVPTCCGSGIA 424
 DB 284 RTVMVTRDFTQVDSYAFESM-----IAAKMETAS-----ELGREED 323
 QY 425 WYNLE-----EERMA-----MLQDPH 441
 DB 324 DVDLELRARFEQLISRRRLNLSVLLRQNP 355
 RESULT 15
 ID POLG HPAVH STANDARD; PRT; 2227 AA.
 AC P08617; P06443; Q81082;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last annotation update)
 DE Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
 DE P2A TO P2C; Probable proteins P3A TO P3C; RNA-directed RNA polymerase
 DE P3D (EC 2.7.7.48)].
 OS Hepatitis A virus (strain HM-175).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Hepatitis A virus.
 OX NCBI_TaxID=12098;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wild type;
 RX MEDLINE=87061253; PubMed=3023706;
 RA Cohen J.I., Ticehurst J.R., Purcell R.H., Buckler-White A.,
 RA Baroudy B.M.;
 RA "Complete nucleotide sequence of wild-type hepatitis A virus:
 RA comparison with different strains of hepatitis A virus and other
 RA picornaviruses".
 RL J. Virol. 61:50-59 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Attenuated;
 RX MEDLINE=87175701; PubMed=3031686;
 RA Cohen J.I., Rosenblum B., Ticehurst J.R., Daemer R.J., Feinstone S.M.,
 RA Purcell R.H.;
 RA "Complete nucleotide sequence of an attenuated hepatitis A virus:
 RA comparison with wild-type virus".
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2497-2501 (1987).
 RN [3]
 RP SEQUENCE OF 1-854 AND 1724-2227 FROM N.A.
 RX MEDLINE=85166289; PubMed=2984684;
 RA Baroudy B.M., Ticehurst J.R., Miele T.A., Maizel J.V. Jr.,
 RA Purcell R.H., Feinstone S.M.;
 RA "Sequence analysis of hepatitis A virus cDNA coding for capsid
 RA proteins and RNA polymerase".
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2143-2147 (1985).
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)_n.
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
 CC each of which is composed of one copy each of proteins VP1, VP2,
 CC VP3, and VP4.
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -!- MISCELLANEOUS: STRAIN HM-175/7 MK-5 IS ATTENUATED STRAIN DERIVED
 CC FROM HM-175 BY 32 PASSAGES IN AFRICAN GREEN MONKEY KIDNEY CELLS.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -!- CAUTION: REF.3 SEQUENCE HAS A NUMBER OF CONFLICTS WITH THAT

CC CC SHOWN.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M14114; AAA45475.1; -

DR EMBL; M14707; AAA45465.1; -

DR EMBL; M14707; AAA45466.1; ALT_INIT.

DR EMBL; M16322; AAA45471.1; -

DR PIR; A03905; A03905.

DR PIR; A25981; GNNYHM.

DR PIR; A94149; GNNYMK.

DR PDB; IHAV; 23-DEC-96.

DR MEROPS; C03.005; -

DR InterPro; IPR004004; Calici_pol_hel.

DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR006005; RNA_helicase.

DR InterPro; IPR007095; RNA_pol_DS_P8.

DR InterPro; IPR001205; RNA_pol_P3D.

DR InterPro; IPR007094; RNA_pol_P5vir.

DR InterPro; IPR008975; Viral_cap_coat.

DR Pfam; PF00680; RNA_dep_RNA_pol; 1.

DR Pfam; PF00910; RNA_helicase; 1.

DR PRINTS; PR00918; CALICIVIRUSNS.

DR Polyprotein; Coat protein; Core protein; Transferase; Hydrolase; Thiol protease; 3D-structure.

DR RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.

FT CHAIN 1 23

FT CHAIN 24 245

FT CHAIN 246 491

FT CHAIN 492 836

FT CHAIN 837 980

FT CHAIN 981 1087

FT CHAIN 1088 1422

FT CHAIN 1423 1496

FT CHAIN 1497 1519

FT CHAIN 1520 1738

FT CHAIN 1739 2227

FT VARIANT 77 77

FT VARIANT 764 764

FT VARIANT 821 821

FT VARIANT 1052 1052

FT VARIANT 1062 1062

FT VARIANT 1118 1118

FT VARIANT 1151 1151

FT VARIANT 1163 1163

FT VARIANT 1277 1277

FT VARIANT 1500 1500

FT VARIANT 1805 1805

FT VARIANT 1930 1930

FT SEQUENCE 2227 AA; 251506 MW; 01B225E7AEB740A6 CRC64;

Query Match 4.0%; Score 95.5; DB 1; Length 2227;

Best Local Similarity 22.0%; Pred. No. 23;

Matches 50; Conservative 39; Mismatches 77; Indels 61; Gaps 12;

QY 197 PRDLPKVVYDMEALWFAGKKVATRFVFGVKAISIODEAKSESVPKPSRV----- 249

Db 782 PRSEEDKRFESHIE-----CRKPKLELVGKQRLKYAQEELSNEVLPPPRKMKGLFSQA 837

QY 250 -----HAVTGFLWKLIAASRA-----ITSCTTSTRLSIAQAQVNLK-TRNMME 292

Db 838 KISLFTVEEHEIMKFSWRGVTADTRALRRGFSLAAGRSVWTLMDAGVLTGLRLINDE 897

QY 293 -----TVLDNATGHLFW-----WAQAILLEL-----SHTTPEISDLKLCDLWNL 331

Db 898 KWTMKDDKIVSLIERKFTSNKYSKVNFFPHGMLDLSEIAANSKDFNMSETDLCFLHL 957

QY 332 NGSVKQCN-GDYPETFGK-----KEGYGRMCE---YLD-FQRTWSSM 368

DB 958 N--PKKINLADRMGLSGVQEIKEQGVGLIARCTFLDSIAGTLKSM 1002

Search completed: July 6, 2004, 13:37:16

Job time : 13.6195 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:50 ; Search time 13.8278 Seconds
(without alignments)
2267.788 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKLVYCGVCHSD.....ERVVKDVRPRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:**

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	971	55.2	361	T12571	cinnamyl-alcohol d
2	963	54.8	337	S28045	cinnamyl-alcohol d
3	960	54.6	357	T05625	cinnamyl-alcohol d
4	959	54.6	359	S28043	cinnamyl-alcohol d
5	935.5	53.2	354	S72477	probable cinnamyl-
6	881	50.1	356	S71179	cinnamyl-alcohol d
7	881	50.1	360	T08581	cinnamyl-alcohol d
8	853	48.5	363	T05624	cinnamyl-alcohol d
9	819.5	46.6	375	D84606	cinnamyl-alcohol d
10	802	45.6	376	E84604	alcohol dehydrogen
11	740	42.1	357	S39509	alcohol dehydrogen
12	739	42.0	357	S49443	cinnamyl-alcohol d
13	719	40.9	357	S49444	cinnamyl-alcohol d
14	686	39.0	357	S23526	cinnamyl-alcohol d
15	681	38.7	357	S23525	cinnamyl-alcohol d
16	666	37.9	357	T09141	cinnamyl-alcohol d
17	645	36.7	357	T05413	cinnamyl-alcohol d
18	641	36.5	358	S31572	cinnamyl-alcohol d
19	629	35.8	354	S60242	cinnamyl-alcohol d
20	628	35.7	367	T02990	cinnamyl-alcohol d
21	621	35.3	367	T02767	cinnamyl-alcohol d
22	616.5	35.1	362	S45094	cinnamyl-alcohol d
23	613.5	34.9	362	D87125	alcohol dehydrogen
24	609	34.6	355	AC2768	alcohol dehydrogen
25	609	34.6	368	E97548	alcohol dehydrogen
26	602	34.2	341	S31571	cinnamyl-alcohol d
27	599.5	34.1	346	H70860	alcohol dehydrogen
28	599.5	34.1	346	UC1376	alcohol dehydrogen
29	596.5	33.9	349	C69583	alcohol dehydrogen

30	587.5	33.4	355	2	E96751	alcohol dehydrogen
31	581.5	33.1	353	2	D83361	alcohol dehydrogen
32	571.5	32.5	352	2	G82719	alcohol dehydrogen
33	568.5	32.3	349	2	D82563	alcohol dehydrogen
34	566.5	32.2	348	2	AH2824	alcohol dehydrogen
35	561	31.9	348	2	G82645	alcohol dehydrogen
36	555.5	31.6	349	2	E64759	probable alcohol d
37	551.5	31.4	349	2	C90676	alcohol dehydrogen
38	551.5	31.4	349	2	F85526	alcohol dehydrogen
39	548.5	31.2	349	2	H82643	alcohol dehydrogen
40	504	28.7	348	1	H64657	probable cinnamyl-
41	503	28.6	350	2	A71857	zinc-dependent alc
42	475	27.0	358	2	B81302	probable alcohol d
43	474	27.0	365	2	H71808	zinc-dependent alc
44	415.5	23.6	336	1	S76928	probable aryl alco
45	405.5	23.1	241	2	G97602	hypothetical prote

ALIGNMENTS

RESULT 1
T12571

C:Species: cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3 - common ice plant

C:Species: Mesembryanthemum crystallinum (common ice plant)

C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 08-Dec-2000

C:Accession: T12571

R:Michalowski, C.B.; Bohnert, H.J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z17531

A:Accession: T12571

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-361 <MIC>

A:Cross-references: EMBL:U79770; NID:g1724109; PID:g1724110

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: alcohol metabolism; oxidoreductase; zinc

F:32-340/Domain: long-chain alcohol dehydrogenase homology <LAD>

F:51,73,167/Binding site: zing, catalytic (Cys, His, Cys) #status predicted

Query Match

Best Local Similarity 55.4%; Score 971; DB 2; Length 361;

Matches 190; Conservative 52; Mismatches 63; Indels 38; Gaps 6;

Qy	2	TGATDVRFKLVYCGVCHSDHMAKNDWGTSTYPIVPGHELGVGVTEVCCKVKFK-SWRQ	60
Db	36	TGEQDVTFRKLVYCGVCHSDHLYIKENWGNVYPAIPGHEIVGVTEVGNKVNFKVGDKV	95
Qy	61	GRCLWHLGRLPTECNCIHLNENPCNLIQTYGSKYVYDGTMTYGVYNNMWTDEHFIYRIP	120
Db	96	GVGCVGSCR-SCSCENHLENCPKMLTYGSTYYDGLTYGYSYDMVVEEHFAVRIP	154
Qy	121	DNPLDGAAPLLCAGITYSPWYYGLDKPGMHLGVENPRFRPRPPLNLFGLWGRSQS	180
Db	155	DNMALDATAPLLCAGVTYSPKHFELDKPGLHGV-----VGLGG-----	195
Qy	181	LVPLLEKGGSYG-----TSPALHES-----LRTDQDQMEANSTMDGIDT	223
Db	196	LGNVAVKFGAKGVTVISTSPNKKDDEAVNRLGADSFVVSREPEQMSAMGLDIDT	255
Qy	224	VPVVRPLEPLISLITKNGKVVTVGVIAVQPLDLPVPLLIIGRQWVAGSAIGMKETQEMID	283
Db	256	VSAHPPLPLLLGLLKSQKRMVGVDPKPLEPFLPQGRKILANGSCIGMKETQEMID	315
Qy	284	FAAEHNITADIEVIPIDYLNAMERVVKDVRPRFVIDVENTL	326
Db	316	FAAKHDIKSDIEVPMYDVTAMERLLKGDVRYRFDIVDANTL	358

RESULT 2
S28045

C:Species: cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3 - parsley (fragment)
C:Species: Petroselinum Crispum (parsley)

C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
 C/Accession: S28045
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.
 EMBO J. 11, 4677-4684, 1992
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara
 A/Reference number: S28043; MUID:93099840; PMID:1464303
 A/Accession: S28045
 A/Molecule type: mRNA
 A/Residues: 1-337 <KIE>
 A/Cross-references: EMBL:X67817; NID:g20446; PIDN:CAA48028.1; PID:g836638
 A/Note: the authors did not translate the codon for residue 337
 C/Genetics:
 A/Genes: ELI3
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F:12-320/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:27,49,143/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.8%; Score 963; DB 2; Length 337;
 Best Local Similarity 55.2%; Pred. No. 1.3e-72;
 Matches 191; Conservative 46; Mismatches 65; Indels 44; Gaps 5;

QY 2 TGATDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHILGVGVTVGVCKVKFKF-----S6
 Db 12 TGDNDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHILGVGVTVGVCKVKFKF-----S6
 QY 57 --SWRQGRCLWGLRPTCEHCILHLENYCPNLIQTYGKYVDGTYGYSNNMVTDEH 114
 Db 72 GVGCLVGSCL-----SCNCDSDSNNCAKQVYATYVDSGTYGYADSMVADQH 124
 QY 115 FIVRPEMLPDLGAPLLCAGITTYSPWRYVGLDKPGHILGVGVTVGVCKVKFKF-----S6
 Db 125 FIVRPEMLPDLGAPLLCAGITTYSPWRYVGLDKPGHILGVGVTVGVCKVKFKF-----S6
 QY 175 GSRLQSLVPLKKEGSGYTSALMHS-----LIRTDQDQMEAMSTMDGI 220
 Db 171 G--LGHVAVKAKAFGANVTVISTESKQKQALEKLGAEFLVSSDSQMQATGFLHG 228
 QY 221 IDTVPAVRPLELISLLKTNKGVTVGVIAVQPLDLPVFPFLIIGRKMVAGSAIGMKETQE 280
 Db 229 IDTVSALHPVPLGLLKVNGKLVWVGAPEKLELFPVFPFLIIGRKMVAGSAIGMKETQE 288
 QY 281 MIDFAEHNTADIEVIPIDYLYNTAMERVKVVDVFRFVIDVENTL 326
 Db 289 MIDFAAQHNITADVEVIPDYVYNTAMERLVKSDVRYRFFVIDVANTI 334

RESULT 3
 T05625
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3-1 - Arabidopsis thaliana
 N/Alternate names: protein F20D10.100
 C/Species: Arabidopsis thaliana (mouse-ear cross)
 C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 15-Oct-1999
 C/Accession: S28044
 R/Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
 submitted to the Protein Sequence Database, February 1999
 A/Reference number: Z15420
 A/Cross-references: EMBL:AL035538
 A/Molecule type: DNA
 A/Residues: 1-357 <BEV>
 A/Cross-references: EMBL:AL035538
 A/Experimental source: cultivar Columbia; BAC clone F20D10
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.
 EMBO J. 11, 4677-4684, 1992
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara
 A/Reference number: S28043; MUID:93099840; PMID:1464303
 A/Accession: S28044
 A/Molecule type: mRNA
 A/Residues: 1-5, 'Q', 7-16, 'N', 18-19, 'V', 21-357 <KIE>
 A/Cross-references: EMBL:X67816; NID:g16266; PIDN:CAA48027.1; PID:g16267
 C/Genetics:
 A/Genes: ELI3-1
 A/Map position: 4

A/Introns: 29/2; 67/2; 290/1
 A/Note: F20D10.100
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:46,68,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.6%; Score 960; DB 2; Length 357;
 Best Local Similarity 55.4%; Pred. No. 2.4e-72;
 Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;

QY 2 TGATDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHILGVGVTVGVCKVKFKSM-RQ 60
 Db 31 TGEKDVRFKLYCGVCHSDHMAKNDWGTSTYPIVPGHILGVGVTVGVCKVKFKSM-RQ 60
 QY 61 GRCLHGRRLPTCEHCILHLENYCPNLIQTYGKYVDGTYGYSNNMVTDEHFIIRIP 120
 Db 91 GVGYNAGSCR-SCDSCNDSDENYCPNLIQTYGKYVDGTYGYSNNMVTDEHFIIRIP 149
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGHILGVGVTVGVCKVKFKSM-RQ 180
 Db 150 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGHILGVGVTVGVCKVKFKSM-RQ 190
 QY 181 LVPLIKEGSGYTSALMHS-----LIRTDQDQMEAMSTMDGI 223
 Db 191 LGHVAVKAKAFGANVTVISTESKQKQALEKLGAEFLVSSDPKQMDAMTMDGI 250
 QY 224 VPAVRPLELISLLKTNKGVTVGVIAVQPLDLPVFPFLIIGRKMVAGSAIGMKETQEMID 283
 Db 251 VPAHPLLELLGLLKVNGKLVWVGAPEKLELFPVFPFLIIGRKMVAGSAIGMKETQEMID 313
 QY 284 FAEHNTADIEVIPIDYLYNTAMERVKVVDVFRFVIDVENTL 326
 Db 311 LAGKNITADIEVIPDYVYNTAMERLVKSDVRYRFFVIDVANTM 353

RESULT 4
 S28043
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) ELI3-2 - Arabidopsis thaliana
 N/Alternate names: protein F20D10.110
 C/Species: Arabidopsis thaliana (mouse-ear cross)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
 C/Accession: S28043; T05626
 R/Kiedrowski, S.; Kwallack, P.; Hahlbrock, K.; Somssich, I.E.; Dangl, J.L.
 EMBO J. 11, 4677-4684, 1992
 A/Title: Rapid activation of a novel plant defense gene is strictly dependent on the Ara
 A/Reference number: S28043; MUID:93099840; PMID:1464303
 A/Accession: S28043
 A/Molecule type: mRNA
 A/Residues: 1-359 <KIE>
 A/Cross-references: EMBL:X67815; NID:g16268; PIDN:CAA48026.1; PID:g16269
 R/Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
 submitted to the Protein Sequence Database, February 1999
 A/Reference number: Z15420
 A/Cross-references: EMBL:AL035538
 A/Molecule type: DNA
 A/Residues: 1-359 <BEV>
 A/Cross-references: EMBL:AL035538
 A/Experimental source: cultivar Columbia; BAC clone F20D10
 C/Genetics:
 A/Genes: ELI3-2
 A/Map position: 4
 A/Introns: 29/2; 67/2; 290/1
 A/Note: F20D10.110
 C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C/Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:46,68,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 54.6%; Score 959; DB 1; Length 359;
 Best Local Similarity 55.1%; Pred. No. 3e-72;
 Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;

```
QY 1 ETGATDVFVKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVGVTEVCVKVKKFMSRQ 60
Db 30 ETGEKDVFKVLFQGLCHSDLHMVNEGMSYFPLVPGHEIVGVTEVGAQVTKATSEK 89
QY 61 -----GRCLWHLRPTCNCIHLENYCPNLQTYGSKYDGTMTYGGYSNNMVTDE 113
Db 90 VGVGLVSSC-----GSCDSCBGMENYCPKSIQTYGFPYDNTITYGYSDEMVCEE 142
QY 114 HFIVRIPONLPLDGAAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPOL 173
Db 143 GFVIRIPONLPLDGAAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPOL 188
QY 174 WGRSLQSLVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAMST 216
Db 189 GG-----LGHVGVKFAKAGTKVTVISTSEKRDENALNGLDADFVSGDPKQIKDAMGT 243
QY 217 MDGIIDTPVAVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVFPPLIGRWVAGSAGGMK 276
Db 244 MDGIIDTPVAVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVFPPLIGRWVAGSAGGMK 303
QY 277 ETQEMIDFAAEHNTADIEVIPIDVNTAMERVVKKQVRFVDFIDVENTL 326
Db 304 ETQEMIDWAGKHNTADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 353
RESULT 5
S72477
probable cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) - tomato
N;Alternate names: mannitol dehydrogenase [misidentification]
C;Species: Lycopersicon esculentum (tomato)
C;Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 04-Feb-2000
C;Accession: S72477
R;Lauter, F.R.
Mol. Gen. Genet. 252, 751-754, 1996
A;Title: Root-specific expression of the LeRee-1 gene in tomato is induced by exposure
to mannitol
A;Reference number: S72477; PMID:97074893; PMID:8917320
A;Accession: S72477
A;Molecule type: DNA
A;Residues: 1-354 <LAU>
A;Cross-references: EMBL:X92855
A;Experimental source: root
C;Comment: the expression of this mannitol dehydrogenase is root-specific and inducible
C;Genetics:
A;Gene: Rse-1
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F;31-337/Domain: long-chain alcohol dehydrogenase homology <LAU>
F;182-211/Region: beta-alpha-beta NADP nucleotide-binding fold
F;47,69,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;100,103,106,114/Binding site: zinc, noncatalytic (Cys) #status predicted
Query Match 53.2%; Score 935.5; DB 2; Length 354;
Best Local Similarity 54.5%; Pred. No. 2.7e-70;
Matches 186; Conservative 53; Mismatches 65; Indels 37; Gaps 6;
QY 2 TGATDVFVKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVGVTEVCVKVKKFSW-RQ 60
Db 32 TGEKDVQFKVYCGVCHSDLHQINNSNSLYPVPVGVTEVGVTEVGVTEVGVTEVGVTEV 91
QY 61 GRCLWHLRPTCNCIHLENYCPNLQTYGSKYDGTMTYGGYSNNMVTDEHFIVRIP 120
Db 92 GVGCLVGSCL- KCNCNDNLENYCPDQIMTVNGVTVGTPTPTGYGSDIMVINEHYVHW 150
QY 121 DNLPDGAAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPCLNGSLQS 180
Db 151 ENLRME-AAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPCLNGSLQS 190
QY 181 LVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGLDITVP 225
Db 191 LGHMAVFAKAGTKVTVISTSEKRDENALNGLDADFVSGDPKQIKDAMSTMDGLDITVP 250
QY 226 AVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVFPPLIGRWVAGSAGGMKQEMIDFA 285
Db 285 AVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVFPPLIGRWVAGSAGGMKQEMIDFA 285
```

```
Db 251 AIHPLPLINLLKTHGKLVWVGAPEKPELPLVPFLLGGKLVAGSAGGMKQEMIDFA 310
QY 286 AEHNTADIEVIPIDVNTAMERVVKKQVRFVDFIDVENTL 326
Db 311 GRKNTDIPGVVMDYVNTALERLLKSDVKYRFLDIGNTL 351
RESULT 6
S71179
cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD1 - Arabidopsis thaliana (fragment)
N;Alternate names: protein T22F8.230
C;Species: Arabidopsis thaliana (mouse-ear cross)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Jul-1999
C;Accession: S71179
R;Somers, D.A.; Nourse, J.P.; Manners, J.M.; Abrahams, S.L.; Watson, J.M.
submitted to the EMBL Data Library, December 1994
A;Description: Sequence of a partial cDNA encoding Arabidopsis thaliana cinnamyl alcohol
dehydrogenase
A;Reference number: S71179
A;Accession: S71179
A;Molecule type: mRNA
A;Residues: 1-356 <SOM>
A;Cross-references: EMBL:L37884; NID:G598068; PID:AAA74746.1; PID:G598069
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C;Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F;31-339/Domain: long-chain alcohol dehydrogenase homology <LAU>
F;182-211/Region: beta-alpha-beta NADP nucleotide-binding fold
F;46,68,162/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;99,102,105,113/Binding site: zinc, noncatalytic (Cys) #status predicted
Query Match 50.1%; Score 881; DB 2; Length 356;
Best Local Similarity 49.1%; Pred. No. 9.4e-66;
Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;
QY 1 ETGATDVFVKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHELGVGVTEVCVKVKKFSNRQ 60
Db 30 DNGENDVTKILFCGCHTDLETKNDWGSYYPVPGHEIVGVIATKVGKNTVKRGGDR 89
QY 61 -----GRCLWHLRPTCNCIHLENYCPNLQTYGSKYDGTMTYGGYSNNMVTDE 113
Db 90 VGVGVISGC-----QSCSCDQDLENYCPQMSFTYNAIGSDGTKNYGGYSNNMVTDE 142
QY 114 HFIVRIPONLPLDGAAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPOL 173
Db 143 RFVLRFPENLPSDGAAPLLCAGITWYSPWRYVGLDKPGMHLGVWPRFRSRPPLNLPOL 188
QY 174 WGRSLQSLVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAMST 216
Db 189 GG-----LGHVAVKIGKAFGLKVTYISSSTKAEAINELGADSLFTVTPQMKAAIGT 243
QY 217 MDGIIDTPVAVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVFPPLIGRWVAGSAGGMK 276
Db 244 MDYIIDIISAVHLYPLGLKLVKGLALGPEKPELPLVGRWVGGSDVGGMK 303
QY 277 ETQEMIDFAAEHNTADIEVIPIDVNTAMERVVKKQVRFVDFIDVENTL 326
Db 304 ETQEMIDFCAKHNTADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 353
RESULT 7
T08581
cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD1 - Arabidopsis thaliana
N;Alternate names: protein T22F8.230
C;Species: Arabidopsis thaliana (mouse-ear cross)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C;Accession: T08581
R;Bevan, M.; Zimmermann, W.; Gruenewald, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Ma-
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16442
A;Accession: T08581
A;Molecule type: DNA
A;Residues: 1-360 <BEV>
A;Cross-references: EMBL:AL050351
A;Experimental source: cultivar Columbia; BAC clone T22F8
C;Genetics:
```

A:Map position: 4
 A:Introns: 33/2; 71/2; 242/3; 294/1
 A:Note: F22F8.230
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F:31-339/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:186-215/Region: beta-alpha-beta NADP nucleotide-binding fold
 F:50,72,166/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 F:103,106,109,117/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 50.1%; Score 881; DB 2; Length 360;
 Best Local Similarity 49.1%; Pred. No. 9.5e-86;
 Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 60
 DB 34 DNGENDVTVKILFCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 93

QY 61 -----GRCWLHGRRLPTCENCIIHLENYCPNLIOTYGSKYDGTMTYGGYNNMTDE 113
 DB 94 VGVGVLISSC-----QSCSCDQDLENYCPQMSFTYNAIGSDGKYNVGGYSENIVDQ 146

QY 114 HPIVRIIDPLDGAAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPLNPLGL 173
 DB 147 RFLVRLPPLDPSGAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPLNPLGL 192

QY 174 WGRSLQSLVPLPKKEGSGYTSALMHS-----LLRTDQDQMEAMST 216
 DB 193 GG-----LGHVAVKIGKAFGLKVTYSSSTKAEAINHLGADSLVITDPOKKAAGT 247

QY 217 MDGIIDTVPVAPRPLEPLISLLKTKNGVVTGIAVQPLDLPVFLIIGRKNVAGSAIGMK 275
 DB 248 MDYIIDTISAVHALYPLGLLKVNGKLIAGLPEKLEPMLPVLGRKNVAGSAIGMK 307

QY 277 ETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVENTL 326
 DB 308 ETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVANS 357

RESULT 8
 T05624
 cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) LCADa - Arabidopsis thaliana
 N:Alternate names: protein F20D10.90
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-Jan-2000
 C:Accession: T05624
 R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15420
 A:Accession: T05624
 A:Molecule type: DNA
 A:Residues: 1-363 <BEV>
 A:Cross-references: EMBL:AL035538
 A:Experimental source: cultivar Columbia; BAC clone F20D10
 C:Genetics:
 A:Gene: LCADa
 A:Map position: 4
 A:Introns: 34/2; 72/2; 169/1; 243/3; 295/1
 A:Note: F20D10.90
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
 F:32-340/Domain: long-chain alcohol dehydrogenase homology <LAD>
 F:187-216/Region: beta-alpha-beta NADP nucleotide-binding fold
 F:51,73,167/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 F:104,107,110,118/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 48.5%; Score 853; DB 2; Length 363;
 Best Local Similarity 47.9%; Pred. No. 2.1e-63;
 Matches 171; Conservative 50; Mismatches 72; Indels 64; Gaps 5;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 60
 DB 35 KTGBEVRKVLVCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 94

QY 57 -----SWRQRCWMLHGLRLPTCENCIIHLENYCPNLIOTYGSKYDGTMTYGGYNN 108
 DB 95 VGVGVLISSC-----TCSCREDQENYCTKAIATYNGVHHDGTINNGYSDH 142

QY 109 MVTDEHPIVRIIDPLDGAAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPL 168
 DB 143 IVDYERAVKIPHTLPVNSAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPL 191

QY 169 NPLGLW-----GSRQLQSLVPLPKKEGSGYTSALMHS-----LLRTDQDQ 209
 DB 192 GLGGLGHIGVRFPAK-----AFGKVTVVSSITGKSKALDNLGADGFLVSTDEQ 241

QY 210 MEAMSTMDGIIDTVPVAPRPLEPLISLLKTKNGVVTGIAVQPLDLPVFLIIGRKNVAG 269
 DB 242 MKAAAGTMDGIIDTVPVAPRPLEPLISLLKTKNGVVTGIAVQPLDLPVFLIIGRKNVAG 301

QY 270 SAIGCMKETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVENTL 326
 DB 302 SGIGGMQETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVENTL 358

RESULT 9
 D84606
 alcohol dehydrogenase (EC 1.1.1.1) [similarity] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Aug-2002
 C:Accession: D84606
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I.
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MIM:20083487; PMID:10617157
 A:Accession: D84606
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-375 <STO>
 A:Cross-references: GB:AE002093; MID:g4417281; PIDN:AAD20406.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g21890
 A:Map position: 2
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: oxidoreductase; zinc
 F:44,66,150/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 46.6%; Score 819.5; DB 2; Length 375;
 Best Local Similarity 47.6%; Pred. No. 1.3e-60;
 Matches 167; Conservative 48; Mismatches 85; Indels 51; Gaps 6;

QY 1 ETGATDVRKVLVCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 60
 DB 28 ENGENDVTVKILFCGVCHSDIHMKNKDWGTSTPIVPGHVLGVVTEGVCKVKFSWRQ 87

QY 61 -----GRCWLHGRRLPTCENCIIHLENYCPNLIOTYGSKYDGTMTYGGYNNMTDE 113
 DB 88 VGVGVLISSC-----QSCSCDQDLENYCPQMSFTYNAIGSDGKYNVGGYSENIVDQ 140

QY 114 HPIVRIIDPLDGAAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPLNPLGL 172
 DB 141 RFLVRLPPLDPSGAPLLCAGITTYSPWRYGLDKPGMHGLGVWPRFRSRPPLNPLGL 186

QY 173 WGRSLQSLVPLPKKEGSGYTSALMHS-----LLRTDQDQMEAMST 215
 DB 187 LGG-----LGHVAVKIGKAFGLKVTYSSSTKAEAINHLGADSLVITDPOKKAAGT 241

QY 216 TMDGIIDTVPVAPRPLEPLISLLKTKNGVVTGIAVQPLDLPVFLIIGRKNVAGSAIGMK 275
 DB 242 TMDGIIDTVPVAPRPLEPLISLLKTKNGVVTGIAVQPLDLPVFLIIGRKNVAGSAIGMK 301

QY 276 KETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVENTL 326
 DB 302 KETQEMIDFAEHNITADIEVIPDYINTAMERVVKGVKQVFRFVIDVENTL 352

```
RESULT 10
E84604
alcohol dehydrogenase (EC 1.-.-.-) [similarity] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Aug-2002
C:Accession: E84604
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <STO>
A:Cross-references: GB:AE02093; NID:g4417268; PIDN:AA020393.1; GSPDB:GNO0139
C:Genetics:
A:Map position: 2
A:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase; zinc
F:44,66,161/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 45.6%; Score 802; DB 2; Length 376;
Best Local Similarity 47.2%; Pred. No. 3.9e-59;
Matches 166; Conservative 47; Mismatches 87; Indels 52; Gaps 7;

QY 1 ETGATDVRFKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHVLGVVTEVGVCKVKFKSWRQ 60
DB 28 ENGENDVTVKLFVCGVCHSDLHTIKNHGFRPIPIPGHEIVGATKVGKVKFKRQGR 87
QY 61 -----GRCWLHGRLEPTCENCIHLENYCPNLIQTYGSKYDGT-MTYGYSNNMVT 112
DB 88 VGVGVIIIGSC-----QSCSCNQDLENYCPKVFTYNSRSSDGTSGNQGYSVDIIV 140
QY 113 EHFIVRPDNLPLDGAAPLLCAGITTYSPWRYGLDK-PGMHLGVWPRFRFRPLNLP 171
DB 141 HRFVLSIPDGLSPSGAPLLCAGITTYSPMKYGYGTMKESGRKLVN-----186
QY 172 GLWGRSLQSLVPELIKESGYGTSPALMHS-----LLRTDQDQMEAM 214
DB 187 GLGG-----LGHIAVKIGKAGLRVTVISRSSEKEREALDGLADSLFVTTDSQCKEAV 241
QY 215 STMDGIDTVAVRPLPLSLKXNGKVTVGIAVQPLDLPLVPLPIIGRKNVAGSITGG 274
DB 242 GTMDFIIDTVSAEHALLPLFSLKXNGKLVAGLPEKPLDPLFSLVLRKXWVGSGQIGG 301
QY 275 MKETQEMIDFAAENITADIEVPIIDYLTAMERVVKDVRFRFVIDVNTL 326
DB 302 MKETQEMLEFCAKHKIVSDIBLIKMSDINSAMDRLAKSDVRFRFVIDVANSL 353

RESULT 11
S39509
cinamyl-alcohol dehydrogenase (EC 1.1.1.195) - Norway spruce
C:Species: Picea abies (Norway spruce)
C>Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 15-Oct-1999
C:Accession: S39509
R:Galliano, H.; Cabane, M.; Eckerskorn, C.; Lottespeich, F.; Sandermann Jr., H.; Ernst, D.
Plant Mol. Biol. 23, 145-156, 1993
A:Title: Molecular cloning, sequence analysis and elicitor-/ozone-induced accumulation o
A:Reference number: S39509; MUID:94033286; PMID:8219046
A:Accession: S39509
A:Molecule type: mRNA
A:Residues: 1-357 <GAL>
A:Cross-references: EMBL:X72675; NID:g393442; PIDN:CAA51226.1; PID:g393443
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
```

```
Query Match 42.1%; Score 740; DB 2; Length 357;
Best Local Similarity 42.7%; Pred. No. 5.4e-54;
Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHVLGVVTEVGVCKVKFKSWRQ-- 60
DB 33 GPEDVIVKVIYCGICHSDLVQMRNEMGMSNYPMVPGHEVGVVTEIGSEVKKFKVGEHV 92
QY 61 -----GRCWLHGRLEPTCENCIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEHF 115
DB 93 VGVGVIIIGSC-----RSCNCNGSMEQYCSKEIWTYNDVNDHGTPTQGGFASWVVDQMF 145
QY 116 IVRIPDNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRFRPLNLPGLWG 175
DB 146 VVRIPENLPLEQAAPLLCAGVTYVSPMKFGMTBPKKCGI-----LGLGVGH 194
QY 176 SRLQSLVPLLIKESGYGTSPALMHS-----LLRTDQDQMEAMSTMD 218
DB 195 MG-----VKIAKAFGLHVTVISSDKKKEALEVLGADAYLVSKQAKQOEARSJD 246
QY 219 GIIDTVPAVRPLPLSLKXNGKVTVGIAVQPLDLPLVPLPIIGRKNVAGSITGGMET 278
DB 247 YIMDTIPVAHPLPLALKTNGKLVMLGVVPEPLHFVTELLILGRSITAGSFIGSMET 306
QY 279 QEMIDFAAENITADIEVPIIDYLTAMERVVKDVRFRFVIDV 322
DB 307 QETLDFCAEKVYSMIEVVGLDYINTAMERLVKNDVRYRFVVDV 350

RESULT 12
S49443
cinamyl-alcohol dehydrogenase (EC 1.1.1.195) A - loblolly pine
C:Species: Pinus taeda (loblolly pine)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Dec-2000
C:Accession: S49443; S55285; S55284
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.
submitted to the EMBL Data Library, September 1994
A:Description: Genetic analysis of Cad in loblolly pine. Single gene inheritance, molec
A:Reference number: S49443
A:Accession: S49443
A:Molecule type: mRNA
A:Residues: 1-357 <MAC>
A:Cross-references: EMBL:Z37991; NID:9558384; PIDN:CAA86072.1; PID:g558385
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.R.; O'Malley, D.M.
Mol. Gen. Genet. 247, 537-545, 1995
A:Title: Genetic analysis of cinamyl alcohol dehydrogenase in loblolly pine: single ge
A:Reference number: S55285; MUID:95327049; PMID:7603432
A:Accession: S55285
A:Molecule type: mRNA
A:Residues: 63-123;189-198;333-357 <MAW>
A:Cross-references: EMBL:Z37991
C:Genetics:
A:Introns: 123/3; 198/3
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 42.0%; Score 739; DB 2; Length 357;
Best Local Similarity 43.5%; Pred. No. 6.5e-54;
Matches 147; Conservative 63; Mismatches 90; Indels 38; Gaps 5;

QY 3 GATDVRFKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHVLGVVTEVGVCKVKFKSWRQ-- 60
DB 33 GPEDVIVKVIYCGICHSDLVQMRNEMGMSHYPMVPGHEVGVVTEIGSEVKKFKVGEHV 92
QY 61 -----GRCWLHGRLEPTCENCIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEHF 115
DB 93 VGVGVIIIGSC-----RSCNCNGSMEQYCSKEIWTYNDVNDHGTPTQGGFASWVVDQMF 145
QY 116 IVRIPDNLPLDGAAPLLCAGITTYSPWRYGLDKPGM-----HLGVWPRFRFRS 164
```



```
Db 146 VVRIPENLPLEQAAPLLCAGVTFS PMKFAMTEBPKKCGILGLGVGHGKIAKAF-- 203
Qy 165 RPPLNPLCLWGRSLQSLVPLPKEGSGVTGSPALMHSILRTDQDMRAAMSTMGIIDTV 224
Db 204 --GHHVTVISSDKKKEEMEVILGADAY-----LVSKDTRKMERABSLDYIMDTI 252
Qy 225 PAVRPLEPLISLLKTKNGKVVTVGIAVQDLPLVPLIIGRKQWAGSAIGGMKETOEMIDF 284
Db 253 PVAHPLEPYLALLTKNGKLVMLGVVPPELHVTPPLIILGRSIAGSFIGGMEETQETLDF 312
Qy 285 AAENHITADIEVIPIDYLTAMERVVKDVRFRFVIDV 322
Db 313 CAEKVSSMIEVGLDYINTAMERLEKNDVRFVVDV 350

RESULT 13
S49444
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) B - loblolly pine
C:Species: Pinus taeda (loblolly pine)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Oct-1999
C:Accession: S49444; S5286; S5295
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.
submitted to the EMBL Data Library, September 1994
A:Description: Genetic analysis of Cad in loblolly pine. Single gene inheritance, molecu
A:Reference number: S49443
A:Accession: S49444
A:Molecule type: mRNA
A:Residues: 1-357 <MAC>
A:Cross-references: EMBL:Z37992; NID:9553386; PIDN:CAA86073.1; PID:9550387
R:MacKay, J.J.; Liu, W.; Whetten, R.; Sederoff, R.; O'Malley, D.M.
Mol. Gen. Genet. 247, 537-545, 1995
A:Title: Genetic analysis of cinamyl alcohol dehydrogenase in loblolly pine: single gen
A:Reference number: S55285; MUID:95327049; PMID:7603432
A:Molecule type: mRNA
A:Residues: 63-123; 189-198; 333-334, 'E', 336-357 <MAW>
A:Cross-references: EMBL:Z37992
C:Genetics:
A:Introns: 123/3; 198/3
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 40.9%; Score 719; DB 2; Length 357;
Best Local Similarity 41.6%; Pred. No. 3e-52;
Matches 143; Conservative 63; Mismatches 88; Indels 50; Gaps 5;

Qy 3 GATDVRFKVLYCGVCHSDIHMAKNDGTSTYPIVPGHELGVTVTEVGCKVKPKSMRQ-- 60
Db 33 GPEDVIVKVIYCGICHSDLVQWNGMSHYPMVPGHEVVGIVTEIGSRVKKPKVGEHV 92
Qy 61 -----GRCLMHLGRLPTECNCIHLNENYCPNLQIOTYGSKYVDGTMVGGYNNMVTDEHF 115
Db 93 VGCIVGSC-----RSCGNCNSMEQYCSKRTWYNDVNDHGTPTGGFASNVVDQMF 145
Qy 116 IVRIPNLPDGAAPLLCAGITTYSPWRYGLDKPGMELGVWFRFRSRPPLNLPGLMG 175
Db 146 VVRIPENLPLEQAAPLLCAGVTFS PMKFAMTEBPKKCGI-----LGLGVGH 194
Qy 176 SRLQSLVPLPKEGSGVTGSPALMHS-----LLRTDQDMRAAMSTM 218
Db 195 MG-----VKIAKAPGLHVTVISSDKKKEEMEVILGADAYLVSKDTRKMERABSLD 246
Qy 219 GIIDTVPAVRPLEPLISLLKTKNGKVVTVGIAVQDLPLVPLIIGRKQWAGSAIGGMK 278
Db 247 YIMDTIPVAHPLEPYLALLTKNGKLVMLGVVPPELHVTPPLIILGRSIAGSFIGSMEET 306
Qy 279 QEMIDFAAENHITADIEVIPIDYLTAMERVVKDVRFRFVIDV 322
Db 307 QETLDFCAEKVSSMIEVGLDYINTAMERLEKNDVRFVVDV 350
```

```
RESULT 14
S23526
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD19 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S23526
R:Knight, M.E.; Halpin, C.; Schuch, W.
Plant Mol. Biol. 19, 793-801, 1992
A:Title: Identification and characterisation of cDNA clones encoding cinamyl alcohol de
A:Reference number: S23525; MUID:92353388; PMID:1643282
A:Accession: S23526
A:Molecule type: mRNA
A:Residues: 1-357 <KNI>
A:Cross-references: EMBL:X62344; NID:919840; PIDN:CAA44217.1; PID:919841
C:Genetics:
A:Gene: CAD19
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 39.0%; Score 686; DB 2; Length 357;
Best Local Similarity 43.1%; Pred. No. 1.7e-49;
Matches 146; Conservative 56; Mismatches 99; Indels 38; Gaps 8;

Qy 2 TGATDVRFKVLYCGVCHSDIHMAKNDGTSTYPIVPGHELGVTVTEVGCKVKPKSMRQ 61
Db 32 TGPEDVIVKVIYCGICHSDLVQWNGMSHYPMVPGHEVVGIVTEIGSRVKKPKVGEHV 88
Qy 62 RCLMHLGRLPTECNCIHLNENYCPNLQIOTYGSKYVDGTMVGGYNNMVTDEHFIVR 118
Db 89 DTGVGVLVSCRCNCGCKREIEQYCNKKIWNQNDVTDGKPTGGFASNVVDQMFVK 148
Qy 119 IPNLPDGAAPLLCAGITTYSPWRYGLDKPGMELGVWFRFRSRPPLNLPGLMGSRL 178
Db 149 IPEGMAPEQAAPLLCAGITTYSPFNHFGFNQSGPRGGI-----LGLGVGHMGV 197
Qy 179 QSLVPLPKEGSGVT-----SPALMH-----SLLRDQDMRAAMSTMGIIDTV 224
Db 198 K-----TAKAMGHVTVISSNNKKRQEALEHLGADAYLVSSDIDKMEADSLDIIDTV 252
Qy 225 PAVRPLEPLISLLKTKNGKVVTVGIAVQDLPLVPLIIGRKQWAGSAIGGMKETOEMID 283
Db 253 PVGHELEPLISLLKIDKLLIGVINTPLQF-IQPMVLMGRKSTGFIGSMKETEMLD 311
Qy 284 FAENHITADIEVIPIDYLTAMERVVKDVRFRFVIDV 322
Db 312 FCXKEGVTIQIEIVKMDYINTAMERLEKNDVRFVVDV 350

RESULT 15
S23525
C:cinamyl-alcohol dehydrogenase (EC 1.1.1.195) CAD14 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S23525
R:Knight, M.E.; Halpin, C.; Schuch, W.
Plant Mol. Biol. 19, 793-801, 1992
A:Title: Identification and characterisation of cDNA clones encoding cinamyl alcohol de
A:Reference number: S23525; MUID:92353388; PMID:1643282
A:Accession: S23525
A:Molecule type: mRNA
A:Residues: 1-357 <KNI>
A:Cross-references: EMBL:X62343; NID:919838; PIDN:CAA44216.1; PID:919839
C:Genetics:
A:Gene: CAD14
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; oxidoreductase; zinc
F:32-340/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:14 ; Search time 44.1542 Seconds
(without alignments)
3229.910 Million cell updates/sec

Title: US-09-857-518A-6
Perfect score: 2373
Sequence: 1 MEKIEVINSKTIKPTSTSS.....MAMLEQDPFLALASPKTLLI 452

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rviro:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2373	100.0	452	10	Q9FVFI fragaria an
2	2043	86.1	455	10	Q8GTM5 fragaria ve
3	538.5	22.7	455	10	Q9LU88 arabidopsis
4	538	22.7	456	10	Q8GV04 capsicum ch
5	506	21.3	426	10	Q9FGV4 arabidopsis
6	501	21.1	443	10	Q9FL40 arabidopsis
7	498.5	21.0	443	10	Q9LI71 arabidopsis
8	498	21.0	436	10	Q9FTM1 arabidopsis
9	485	20.4	446	10	Q23392 arabidopsis
10	468.5	19.7	435	10	Q9FYM0 arabidopsis
11	454.5	19.2	428	10	Q9FLW4 arabidopsis
12	450	19.0	99	10	Q23943 arabidopsis
13	448	18.9	435	10	Q23393 arabidopsis
14	438	18.5	474	10	Q94FT4 papaver som
15	406	17.1	443	10	Q8GZU0 catharantha
16	404.5	17.0	433	10	O64988 clarkia bre

17	403.5	17.0	431	10	Q9SQ00	Q9sq00 clarkia con
18	398.5	16.8	433	10	Q9SPU3	Q9spu3 clarkia bre
19	390.5	16.5	431	10	Q9SQ01	Q9sq01 clarkia con
20	387.5	16.3	439	10	Q9ZTK5	Q9ztk5 catharanth
21	384	16.2	430	10	Q9SQ02	Q9sq02 clarkia con
22	318	13.4	328	10	Q9SPZ9	Q9spz9 clarkia con
23	285	12.0	219	10	Q9M5K9	Q9m5k9 euphorbia e
24	262	11.0	430	10	Q9SMW7	Q9smw7 arabidopsis
25	259.5	10.9	449	10	Q7XPK7	Q7xpk7 oryza sativ
26	253	10.7	426	10	Q8LE88	Q8le88 arabidopsis
27	253	10.7	426	10	Q9FFQ7	Q9ffq7 arabidopsis
28	248	10.5	439	10	Q84QG6	Q84qg6 hordeum vul
29	245.5	10.3	435	10	Q8GSM7	Q8gsm7 nicotiana t
30	242	10.2	446	10	Q23917	Q23917 dianthus ca
31	242	10.2	461	10	Q9LE83	Q9le83 arabidopsis
32	234.5	9.9	461	10	Q9LF70	Q9lf70 arabidopsis
33	233.5	9.8	445	10	Q23918	Q23918 dianthus ca
34	232	9.8	431	10	Q9SST8	Q9sst8 ipomoea bat
35	232	9.8	442	10	Q23916	Q23916 dianthus ca
36	231	9.7	424	10	Q9FTG9	Q9ftg9 oryza sativ
37	230	9.7	440	10	Q7XXP2	Q7xxp2 avena sativ
38	228.5	9.6	451	10	O64470	O64470 arabidopsis
39	228	9.6	464	10	Q9MAP9	Q9map9 arabidopsis
40	227.5	9.6	433	10	Q9FI78	Q9fi78 arabidopsis
41	227.5	9.6	433	10	Q8LFT5	Q8lft5 arabidopsis
42	225	9.5	445	10	O24645	O24645 dianthus ca
43	222.5	9.4	441	10	Q9FLM5	Q9flm5 arabidopsis
44	222.5	9.4	457	10	Q94CD1	Q94cd1 arabidopsis
45	222	9.4	572	10	O64549	O64549 arabidopsis

ALIGNMENTS

RESULT 1

Q9FVFI	ID	Q9FVFI	PRELIMINARY;	PRT;	452 AA.
AC	Q9FVFI				
DT	01-MAR-2001	(T-EMBLrel. 16, Created)			
DT	01-MAR-2001	(T-EMBLrel. 16, Last sequence update)			
DT	01-JUN-2003	(T-EMBLrel. 24, Last annotation update)			
DE	Alcohol acyltransferase.				
GN	AAAT.				
OS	Fragaria ananassa (Strawberry).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosids;				
OC	eurosid1; Rosales; Rosaceae; Rosoideae; Fragaria.				
OX	NCBI_TaxID=3747;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fruit;				
RX	MEDLINE=2021765; PubMed=10810141;				
RA	Aharoni A., Keizer L.C.P., Boumeester H.J., Sun Z.,				
RA	Alvarez-Huerta M., Verhoeven H.A., Blaas J.,				
RA	Van Houtwelingen A.M.L., De Vos R.C.H., Van der Voet H.,				
RA	Jansen R.C., Guis M., Mol J., Davis R.W., Schena M., van Tunen A.J.,				
RA	O'Connell A.P.O.;				
RT	Identification of the SAAT gene involved in strawberry flavor				
RT	biogenesis by use of DNA microarrays."				
RL	Plant Cell 12:647-662(2000).				
DR	EMBL; AF193789; AAG13130.1; -				
DR	GO; GO:0008415; F:acyltransferase activity; IEA.				
DR	GO; GO:0016740; F:transferase activity; IEA.				
DR	InterPro; IPR003480; Transferase.				
DR	PFam; PF02459; Transferase; 1.				
KW	Acyltransferase; Transferase.				
SK	SEQUENCE 452 AA; 50701 MW; 3D14F3BA732B135E CRC64;				

Query Match 100.0%; Score 2373; DB 10; Length 452;
Best Local Similarity 100.0%; Pred. No. 8.6e-208;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEKIEVINSKTIKPTSTSTPLQPKTLLDQLTPPAYVPIFFYPTIDHFNLPQTLA 60

QY 4 IEVSINSKHTIKPSTSTPLQ--YKLTLLDQTPPAYVPIVFFYPITDHDNLPQTLD 61
 DB 3 LEITVTSQELVSPSRNHNPPCHHLSLFDQAPLPIFMPFFLFY-----HN---KTNLSD 55
 QY 62 -----LRQALSETLTLYPLSGRVKN--LYIDDFEEGVPYLEARVNCMDTDFLRLRK 112
 DB 56 KERSDHIKSSLSEILNLYPLAGRIKNSGDVVVNCNDV--GVSVFAKADCNMSQILENPN 113
 QY 113 IECLNEFVPIKPSMAISDERYPPLGVQVNP--SGIAIGVSVSHKLDGGTADCFELKS 171
 DB 114 ---PHELNLKHPFEFEVSD--VPLT-VQLTTFECGLGALGLSHKLDALSGLLIFVNS 167
 QY 172 WGAVERGCRNIHPSLSAALLPPRDPOLPKYVDOME--ALWFAKVKVATRRFVGVKA 230
 DB 168 WAAFAFGQTDLEITTSF-DLAKWFPD-----IENLNWAGITENIVTRFVLRSS 220
 QY 231 ISSIOEAKSESVPKPSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAQAQVNLTRMN 290
 DB 221 VESLRETFSGNKIRATRVEVLVSFTWSRFMAST---NHDDKTGKIYTLIHPVNLRRQAD 277
 QY 291 METVLDNATGNLFWAAQAILELSHTTPEI-----SOLKCLDYNLLNGSVKQCGDYFETP 346
 DB 278 PD-IPDNMFGN-----IMRFSVTYPMIINENDEEKASLDVDMREERKIDAVVKKL 329
 QY 347 KCKEGYGRMCEYLDFTWSSMEPADIVLFSSWTNFFNPL---DFGWGRTSMIGVAGKI 403
 DB 330 Q-EDNRG-----HLEFLNKQASGFVNGEIVSFSTLSCKFPVYEADFGWGLNVASA---381
 QY 404 ESASCKFIILVPTQCGSGIEAWNLEEKAMLEQDPHFLALASPKTLI 452
 DB 382 RMSYKNLVAFIDTKEGDGIEAWNLDQNDMSRFEADBEILLRYVSSNPSVMV 432
 RESULT 11
 Q9FLW4 ID Q9FLW4 PRELIMINARY; PRT; 428 AA.
 AC Q9FLW4; STRAIN=Columbia;
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Acetyl-CoA:benzylalcohol acetyltransferase-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98290346; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RA "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned pl and TAC clones.";
 RL DNA Res. 5:41-54 (1998).
 DR EMBL; AB009056; BAB08720.1; -
 DR InterPro; IPR003480; Transferase.
 DR Pfam; PF02458; Transferase; 1.
 SQ SEQUENCE 428 AA; 4788 MW; 6CFF3F61F5703C4C CRC64;

Query Match 19.2%; Score 454.5; DB 10; Length 428;
 Best Local Similarity 28.7%; Pred. No. 1.2e-32;
 Matches 133; Conservative 87; Mismatches 192; Indels 51; Gaps 17;

QY 4 IEVSINSKHTIKPSTSTPLQYKLTLLDQTPPAYVPIVFFYPITDHDNLPQTLD 63
 DB 1 MKLELLSKEVKEPA-SPNHLQTLISLFDQFLSTVSAIFFY---DDHNSQEDIQRLX 55
 QY 64 QALSETLTLYPLSGRVKNLYIDDFEEGVPYLEARVNCMDTDFLR-LRKEICLNRFVPI 122
 DB 56 SSLSQTLSLFLYPLAGQIKGVTGHCNDEGALFTEARAEIFLSDFLRNPSDADLIQKFI-- 113

QY 123 KPFSMAISDERYPPLGVQVNP--DSGIAIGVSVSHKLDGGTADCFELKSWGAVRGRE 181
 DB 114 --VSPHAPETWPLLVKVIPEKDKGFVAVSVSHKICDAASLSFVCSWTASKGYAD 171
 QY 182 NIHPSLSAALLPPRDP--DLPEKYVDOMEALWFAKVKVATRRFVGVKAISISIDEA 238
 DB 172 T-VDPEF-VGADFPADISIEPPLLVHE-----TKSKTKRFVFGSLMIEKLNRA 221
 QY 239 KS-ESVPSKPSRVHVTGFLWKHLIAASRALTSCTTSTRLSIAQAQVNLTRMNMTVLDN 297
 DB 222 SSKRPVQATRIEISITALLRLCMWKAGHSKGVKGEFAIT---QTMOLRPRVSSSLPHK 278
 QY 298 ATGKLFWAQAILELSHTTPEISOL--KLCCLDYNLLNGSVKQCGDYFETFKGKEG---351
 DB 279 AIGN-FFPLPLKESSEKWEIEETYSKLOKQTELINRDNSEDASVFAKERIASA 337
 QY 352 -YGRMCEYLDFTWSSMEPADIVLFSSWTNFFNPLDFGWGRTSMIGVAGKIESASCK 409
 DB 338 MLSSICE-----ISPMEYAVSSWCRMSFYEAENFGWGPVWVAP-----DSVDKT 383
 QY 410 FIILVPTQCGSGIEAWNLEEKAMLEQDPHFLALASPKTLI 452
 DB 384 QVVLMDSKDSGVEARVTLPTDVMKYEHDSSELEVTATPSFSI 426

RESULT 12

O23943 ID O23943 PRELIMINARY; PRT; 99 AA.
 AC O23943; STRAIN=cv. Reine des Vallées;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JUN-1998 (Tremblrel. 05, Last sequence update)
 DE Ripening-induced protein (Fragment).
 OS Fragaria vesca (Woodland strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Rosales; Rosaceae; Rosoideae; Fragaria.
 OX NCBI_TaxID=57918;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Reine des Vallées;
 RA Nam Y.W., Ticht L., Leperlier M., Querq B., Marty I., Lelievre J.M.;
 RA "Isolation and characterization of cDNAs from genes differentially
 RT expressed during ripening of wild strawberry (Fragaria vesca L.).";
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ001450; CAA04771.1; -
 DR InterPro; IPR003480; Transferase.
 DR Pfam; PF02458; Transferase; 1.
 FT NON TER 1
 SQ SEQUENCE 99 AA; 11336 MW; 3B567A4617095858 CRC64;

Query Match 19.0%; Score 450; DB 10; Length 99;
 Best Local Similarity 87.5%; Pred. No. 3.3e-33;
 Matches 84; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 357 EYLDFTWSSMEPADIVLFSSWTNFFNPLDFGWGRTSMIGVAGKIESASCKFIILVPT 416
 DB 1 EYLDFTWSSMEPAEIVLFTSWTFFNQLDFTGWGRTSMIGVAGKIESAFCNLTIVPT 60
 QY 417 QCGSGIEAWNLEEKAMLEQDPHFLALASPKTLI 452
 DB 61 PCDTGIEAWNLEEKAMLEQDPHFLALASPKTLI 96

RESULT 13

O23393 ID O23393 PRELIMINARY; PRT; 435 AA.
 AC O23393; STRAIN=cv. Reine des Vallées;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JUN-1998 (Tremblrel. 05, Last sequence update)
 DE Hypothetical protein (HSR201 like protein).
 GN AT4G15400.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:24 ; Search time 42.3213 Seconds
(without alignments)
2176.455 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ETGATDVRFKVLVYGVCHSD.....ERVVKDVRFRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1758	100.0	326	3 AAB36445	Strawberr
2	1758	100.0	326	3 AAY79662	Strawberr
3	1007	57.3	355	6 ADA38387	Soybean c
4	991	56.4	333	3 AAB36444	Strawberr
5	991	56.4	333	3 AAY79661	Strawberr
6	979	55.7	362	5 AAU80013	Sinapyl a
7	979	55.7	362	5 AAU79807	Sinapyl a
8	979	55.7	362	7 ADD33896	Quaking a
9	960	54.6	357	3 AAG29470	Arabidops
10	960	54.6	357	5 ABB33285	Herbicida
11	959	54.6	359	5 ABB33286	Herbicida
12	881	50.1	360	3 AAG40179	Arabidops
13	881	50.1	360	5 ABB33303	Herbicida
14	876	49.8	360	3 AAG25506	Arabidops
15	871	49.5	306	3 AAG29471	Arabidops
16	869	49.4	360	6 ADA38379	Soybean c
17	832	47.3	371	6 ADA38389	Wheat cin
18	827	47.0	283	3 AAB36447	Strawberr
19	827	47.0	283	3 AAY79664	Strawberr
20	824.5	46.9	278	3 AAB36446	Strawberr
21	824.5	46.9	278	3 AAY79663	Strawberr
22	819.5	46.6	375	3 AAG32129	Arabidops
23	819.5	46.6	375	5 ABB91832	Herbicida
24	802	45.6	376	3 AAG32132	Arabidops
25	800	45.5	361	6 ADA38383	Corn cinn

ALIGNMENTS

RESULT 1

AAB36445
ID AAB36445 standard; protein; 326 AA.

XX AC AAB36445;

XX AC
DT 28-FEB-2001 (first entry)

DE Strawberry alcohol dehydrogenase protein SEQ ID NO:7B.

KW Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;
KW aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;
KW pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;
KW alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;
KW food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;
KW confectionery; flavouring; oral medication; vitamin; aroma; beverage;
KW alcohol; scent; fragrance; perfume; cosmetic; suspension aid;
KW aluminum salt; anti-perspirant; pharmaceutical; cleaning product;
KW insect pheromone; dye carrier; solvent; insect repellent; miticide;
KW scabicide; plasticiser; deodorant.

XX OS Fragaria x ananassa.

XX WC2000032789-A1.

XX PD 08-JUN-2000.

XX PF 02-DEC-1999; 99WO-NL000737.

XX PR 02-DEC-1998; 98EP-00204018.

XX PR 12-MAR-1999; 99EP-00200739.

XX PA (CPRO-) CPRO-DLO CENT PLANTENVERDELINGS REPROD.

XX PI Aharoni A, Luecker J, Vertoeven HA, Van Tunen AJ, O'Connell AP;

XX N-PSDB; 2000-412335/35.

XX WPI; 2000-412335/35.

PT A new DNA sequence encoding a polypeptide with alcohol acyl transferase
PT activity for producing and regulating aromatic and/or aliphatic ester
PT formation in microorganisms, plant cells or plants.

XX Claim 32; Page 89-90; 163pp; English.

XX The present invention describes nucleotide sequences with thiolase,
CC alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,
CC aminotransferase and esterase activities, which are involved in the

CC biosynthetic pathway for aliphatic and/or aromatic ester production in
 CC fruit. The nucleotide sequences can be inserted into the genome of a
 CC fruit-producing plant to regulate aliphatic and/or aromatic ester
 CC formation. Aromatic and/or aliphatic esters in microorganisms, plant
 CC cells or plants are produced by inserting thiolase, alcohol acyl
 CC transferase, alcohol dehydrogenase, pyruvate decarboxylase,
 CC aminotransferase and esterase nucleotide sequences into the genome and
 CC feeding the microorganism or plant with alcohol, aldehydes, alpha-keto
 CC acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and
 CC their proteins can be used in the processed food industry as food
 CC additives to enhance the flavour of syrups, ice-creams, frozen desserts,
 CC yoghurts and confectionery. They are used as flavouring agents for oral
 CC medications and vitamins; provide flavour and aroma in beverages,
 CC including alcohol; enhance or reduce fruit flavour, aroma, fragrance or
 CC scent; enhance the flavour or aroma of natural, synthetic or artificial
 CC products; for the production of novel combinations of artificial flavour
 CC substances; as antibacterial or anti-fungal agents; as fragrance or
 CC perfumes in cosmetics, creams, sun-protectant products, hair
 CC conditioners, lengthening agents and fixatives in perfumes, suspension
 CC aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning
 CC products, personal care products and animal care products; as
 CC disinfectant additives; as degreasing solvents for electronics; as insect
 CC pheromones; and as dye carriers, solvents, insect repellents, miticides,
 CC scabicides, plasticizers and deodorants. The present sequence represents
 CC the specifically claimed strawberry alcohol dehydrogenase
 CC

XX Sequence 326 AA;

Query Match 100.0%; Score 1758; DB 3; Length 326;
 Best Local Similarity 100.0%; Pred. No. 4e-171;
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVVTEVGCCKVFKSWRQ 60
 DB 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVVTEVGCCKVFKSWRQ 60
 QY 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHPIVRIP 120
 DB 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHPIVRIP 120
 QY 121 DNLPDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRFRPPLNLPGLWGSRLQS 180
 DB 121 DNLPDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRFRPPLNLPGLWGSRLQS 180
 QY 181 LVPLIKEGSGSYGTSPLMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240
 DB 181 LVPLIKEGSGSYGTSPLMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240
 QY 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETQEMIDFAAENHITADIEVIPID 300
 DB 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETQEMIDFAAENHITADIEVIPID 300
 QY 301 YLNTAMERVVKQVRFVIDVENTL 326
 DB 301 YLNTAMERVVKQVRFVIDVENTL 326

RESULT 2

AAV79662

ID AAV79662 standard; protein; 326 AA.

XX

AC AAV79662;

XX

DT 12-SEP-2003 (revised)

DT 29-AUG-2000 (first entry)

XX

DE Strawberry alcohol dehydrogenase SLF193 (C-terminal sequence).

XX

KW Strawberry; alcohol dehydrogenase; fruit; ripening; ester; flavour;

KW aroma; transgenic plant.

XX

OS Fragaria x ananassa.

XX

PN BP1006190-A1.

XX 07-JUN-2000.

XX 02-DEC-1998; 98EP-00204018.

XX 02-DEC-1998; 98EP-00204018.

XX (CPRO-) CPRO-DLO CENT PLANTENVERDEDELINGS REPROD.

XX Verhoeven HA, Van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;

PI WPI; 2000-378264/33.

DR N-PSDB; AAA27672.

XX New polynucleotides encoding enzymes from the biosynthetic pathway for

XX aromatic and/or aliphatic ester production in fruit used to modify plant

XX flavors.

XX Claim 36; Page 81-82; 116pp; English.

XX The present sequence is that of the C-terminal region of strawberry cv.

XX Elsanta alcohol dehydrogenase SLF193, an enzyme that shows upregulated

XX expression in ripening fruit. The invention relates to DNA sequences (see

XX AAA27666-78) encoding enzymes (see AAV29656-68) involved in the metabolic

XX pathway leading to the formation of aliphatic and/or aromatic esters in

XX ripening fruit. The enzymes have alcohol acyl transferase, alcohol

XX dehydrogenase, pyruvate decarboxylase, thiolase or aminotransferase

XX activity. Expression vectors comprising the DNA sequences may be used to

XX regulate ester formation in fruit. Genetically modified plants, plant

XX cells and microorganisms can be used to produce esters. The DNA

XX sequences, polypeptides and antibodies are also used to screen fruit: for

XX volatile ester compounds; for quality such as flavour, fragrance, aroma,

XX scent, texture or shape, to distinguish between cultivars and varieties;

XX and to monitor harvest time, post-harvest quality, shelf-life, timing of

XX pesticide application, and resistance capacity based on volatile ester

XX profiles. (Updated on 12-SEP-2003 to standardise CS field)

XX

SQ Sequence 326 AA;

Query Match 100.0%; Score 1758; DB 3; Length 326;

Best Local Similarity 100.0%; Pred. No. 4e-171;

Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVVTEVGCCKVFKSWRQ 60

DB 1 ETGATDVRFKVLYCGVCHSDIHMAKNDWGTSTYPIVPGHGLVGVVTEVGCCKVFKSWRQ 60

QY 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHPIVRIP 120

DB 61 GRCLHGLRLPTCNCIHLENYCPNLIQTYGSKYDGTMTYGGYNNMVTDEHPIVRIP 120

QY 121 DNLPDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRFRPPLNLPGLWGSRLQS 180

DB 121 DNLPDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRFRPPLNLPGLWGSRLQS 180

QY 181 LVPLIKEGSGSYGTSPLMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240

DB 181 LVPLIKEGSGSYGTSPLMHSLLRTDQDQEAAMSTMDGIIDTVPVAVRPLEPLISLKTN 240

QY 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETQEMIDFAAENHITADIEVIPID 300

DB 241 GKVVTVGIAVQPLDLPVFPPLIIGRKWAGSAIGGMKETQEMIDFAAENHITADIEVIPID 300

QY 301 YLNTAMERVVKQVRFVIDVENTL 326

DB 301 YLNTAMERVVKQVRFVIDVENTL 326

RESULT 3

ADA38387

ID ADA38387 standard; protein; 355 AA.

XX

Db 35 TGEDVRFKVLKYGCHSDLSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 94
 QY 61 -----GRCWLHGLRPTCENCIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEH 114
 Db 95 GWGCLVGAC-----HSCESCANDLENYCPKMLTYASIHDTGITYGGYSDHVMANER 147
 QY 115 FIVRIPDNLPDGAAPLLCAGITTYSPWRYGGLDKPGMLGVWPRFRSRPPLNLPGLW 174
 Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGDLDFGKHIGI-----VGLG 193
 QY 175 GSRLOSLVPLIKEGGS-----YGTSPALMSHLRT-----DODQMEAMSTWDGI 220
 Db 194 G--LGHVAVKFAKAFGSKVTIVSTSPSKKEALNFGADSLVSRDQEQMQAAGTLDGI 251
 QY 221 IDTVPAVRPLEPLISLILKNGKVVTVGIAVQPLDLFVPELIIIGRKMVAGSAIGMKETO 280
 Db 252 IDTVSAVHPPLPLFGLKSHGKLLIVGAPEKPLELPAFSLIAGRKIVAGSIGMKETO 311
 QY 281 MIDFAAHEHNITADIEVPIIDYLNLTAMERVVKKVVRFRFVIDVENTL 326
 Db 312 MIDFAAHEHNITADIEVISTDYLNLTATIERLAKNDVRYRFVIDGNTL 357

RESULT 8

ADD93896
 ID ADD93896 standard; protein; 362 AA.

AC ADD93896;

DT 29-JAN-2004 (first entry)

DE Quaking aspen sinapyl alcohol dehydrogenase (SAD).

EX Quaking aspen; Plant; enzyme; 4-coumarate-CoA ligase; 4CL;

KW conferyl aldehyde 5-hydroxylase; CAL35H;

KW SAM-dependent 5-hydroxyconiferaldehyde O-methyltransferase;

KW S-adenosyl-L-Methionine; ALD0MT; coniferyl alcohol dehydrogenase; CAD;

KW sinapyl alcohol dehydrogenase; SAD; transgenic; agronomic property;

KW lignin; cellulose; syringyl/guaiacyl lignin ratio; growth; wood quality;

KW stress resistance; sterility; grain yield; nutritional value; S/G ratio;

KW paper delignification; pulp manufacture; grass digestibility.

XX Populus tremuloides.

XX US2002138870-A1.

XX 26-SEP-2002.

XX 06-MAR-2002; 2002US-00091009.

XX 05-SEP-2000; 2000US-0230086P.

XX 05-SEP-2001; 2001US-00947027.

XX (UNMT) UNIV MICHIGAN TECHNOLOGICAL.

XX Chiang VLC, Li L;

XX WPI; 2003-843045/78.

XX N-PSDB; ADD93895.

PT Genetic transformation of plants, useful for altering lignin and
 PT cellulose contents, by introducing genes from the phenylpropanoid
 PT pathways, also new transgenic plants.

PS Disclosure; SEQ ID NO 2; 55pp; English.

XX The invention relates to the genetic transformation of a plant
 CC simultaneously with several genes from the phenylpropanoid pathways by
 CC incorporating into the genome the genes for 4-coumarate-CoA ligase (4CL),
 CC conferyl aldehyde 5-hydroxylase (CAL35H), S-adenosyl-L-Met-dependent 5-
 CC hydroxyconiferaldehyde O-methyltransferase (ALD0MT), coniferyl alcohol
 CC dehydrogenase (CAD) and sinapyl alcohol dehydrogenase (SAD), their

CC fragments or combination to produce plants with altered agronomic traits.
 CC Also included are preparing plant cells that contain, in the genome,
 CC several DNA constructs containing one or more of the specified genes,
 CC preparing transgenic plants with altered lignin or cellulose contents (by
 CC regenerating plants from the cells and expressing the construct in cells
 CC of the regenerated plant), transgenic plants produced by method above and
 CC their progeny, plants that include in the genome a construct containing
 CC at least one of the specified gene (linked to a promoter and a
 CC terminator) a set of DNA constructs (each containing a promoter,
 CC terminator and at least one of the specified genes or their fragments)
 CC and the set of above or a single construct containing the 4CL gene, where
 CC incorporated into a plant genome. The method is used to transform plants,
 CC particularly trees but also forage crops and monocotyledons, to alter
 CC their agronomic properties, especially lignin and cellulose contents,
 CC syringyl/guaiacyl (S/G) lignin ratio, growth, wood quality, stress
 CC resistance, sterility, grain yield and nutritional value, particularly to
 CC increase S/G ratio (this simplifies delignification in paper and pulp
 CC manufacture) and to increase digestibility of grasses. The present
 CC sequence represents quaking aspen SAD.
 XX
 SQ Sequence 362 AA;

Query Match 55.7%; Score 979; DB 7; Length 362;

Best Local Similarity 56.6%; Pred. No. 3.2e-91;

Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

QY 2 TGATDVRFKVLKYGCHSDLSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 60

Db 35 TGEDVRFKVLKYGCHSDLSIKNDWGFMSYPLVPGHEIVGEVTEVGSVKVNVGDKV 94

QY 61 -----GRCWLHGLRPTCENCIHLENYCPNLIQTYGSKYDGTMTYGGYSNNMVTDEH 114

Db 95 GWGCLVGAC-----HSCESCANDLENYCPKMLTYASIHDTGITYGGYSDHVMANER 147

QY 115 FIVRIPDNLPDGAAPLLCAGITTYSPWRYGGLDKPGMLGVWPRFRSRPPLNLPGLW 174

Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGDLDFGKHIGI-----VGLG 193

QY 175 GSRLOSLVPLIKEGGS-----YGTSPALMSHLRT-----DODQMEAMSTWDGI 220

Db 194 G--LGHVAVKFAKAFGSKVTIVSTSPSKKEALNFGADSLVSRDQEQMQAAGTLDGI 251

QY 221 IDTVPAVRPLEPLISLILKNGKVVTVGIAVQPLDLFVPELIIIGRKMVAGSAIGMKETO 280

Db 252 IDTVSAVHPPLPLFGLKSHGKLLIVGAPEKPLELPAFSLIAGRKIVAGSIGMKETO 311

QY 281 MIDFAAHEHNITADIEVPIIDYLNLTAMERVVKKVVRFRFVIDVENTL 326

Db 312 MIDFAAHEHNITADIEVISTDYLNLTATIERLAKNDVRYRFVIDGNTL 357

RESULT 9

AAG29470

ID AAG29470 standard; protein; 357 AA.

AC AAG29470;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 35071.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147453P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148555P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155119P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.

```
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.
```

```
Query Match 54.6%; Score 960; DB 3; Length 357;
Best Local Similarity 55.4%; Pred. No. 2.8e-89;
Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;
```

```
QY 2 TGATVRFKLVCGVCHSDIHKANDWGSTSTPIVPGHVLGVVTVGCKVKFKSW-RQ 60
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 TGEKDVRFKLVCGICHTDLSMAKNEWGLTTPVPGHEIVGVTVGAKVKKFNAGDKV 90
QY 61 GRCWLHGLRLPTCENCIHLENYCPNLIQTYGSKYDGMTYGGYNNWVTDHFVIRIP 120
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 GVGYNAGSCR-SCDSCNDGDENYCPKMLTSGAKNFDDTWHGGYSDHVCADDFIIRIP 149
QY 121 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGMHGVWEPFRFRPPLMLPGLWSRLQS 180
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 DNLPLDGAAPLLCAGVTVVSPMKYHGLDKPGMHGV-----VGLGG----- 190
QY 181 LVPPLLKEGSGVGTSPALMHS-----LLRTDQDQMEAMSTMDGIIDT 223
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 LGHVAVKFAKMGTKVTYVTSERKRDVAVTLRGADAFVLSRDPKQMDAMGTMDGIIDT 250
QY 224 VPAVRPLEPLISLLKNGKVVTVGIAVQDLPVFPPLIIGRWVAGSAIGGMKETOEMID 283
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 VSATHPLPLLLGLLKNKGLVWVGPAPAELELPVFPPLIFGRKRVVGSWVGGIKETQEMVD 310
QY 284 FFAENITADIEVPIIDYINTAMERVVKQVFRFVIDVENTL 326
DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 LAGKENITADIELISADYVNTAMERLAKADVKYRFVIDVANTM 353
```

```
RESULT 10
ABB93285
ID ABB93285 standard; protein; 357 AA.
XX
XX ABB93285;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2496.
```

```
XX
XW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN W0200210210-A2.
XX
PD 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP009892.
XX
XX 28-AUG-2001; 2001WO-EP009892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
SQ Sequence 357 AA;
```

```
Query Match 54.6%; Score 960; DB 5; Length 357;
Best Local Similarity 55.4%; Pred. No. 2.8e-89;
Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;
```

QY 2 TGATVRFKLVCGVCHSDIHKANDWGSTSTPIVPGHVLGVVTVGCKVKFKSW-RQ 60

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

31 TGEKDVRFKLVCGICHTDLSMAKNEWGLTTPVPGHEIVGVTVGAKVKKFNAGDKV 90

QY 61 GRCWLHGLRLPTCENCIHLENYCPNLIQTYGSKYDGMTYGGYNNWVTDHFVIRIP 120

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

91 GVGYNAGSCR-SCDSCNDGDENYCPKMLTSGAKNFDDTWHGGYSDHVCADDFIIRIP 149

QY 121 DNLPLDGAAPLLCAGITTYSPWRYVGLDKPGMHGVWEPFRFRPPLMLPGLWSRLQS 180

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

150 DNLPLDGAAPLLCAGVTVVSPMKYHGLDKPGMHGV-----VGLGG----- 190

QY 181 LVPPLLKEGSGVGTSPALMHS-----LLRTDQDQMEAMSTMDGIIDT 223

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

191 LGHVAVKFAKMGTKVTYVTSERKRDVAVTLRGADAFVLSRDPKQMDAMGTMDGIIDT 250

QY 224 VPAVRPLEPLISLLKNGKVVTVGIAVQDLPVFPPLIIGRWVAGSAIGGMKETOEMID 283

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

251 VSATHPLPLLLGLLKNKGLVWVGPAPAELELPVFPPLIFGRKRVVGSWVGGIKETQEMVD 310

QY 284 FFAENITADIEVPIIDYINTAMERVVKQVFRFVIDVENTL 326

DQ |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

311 LAGKENITADIELISADYVNTAMERLAKADVKYRFVIDVANTM 353

RESULT 11

ABB93286

ID ABB93286 standard; protein; 359 AA.

XX

XX ABB93286;

XX

```
DT 31-MAY-2002 (first entry)
XX AC
DE Herbicidally active polypeptide SEQ ID NO 2497.
XX KW
XX Herbicidal; plant; agriculture; herbicide.
XX OS
XX Arabidopsis thaliana.
XX FN W0200210210-A2.
XX PD 07-FEB-2002.
XX XX
XX 28-AUG-2001; 2001WO-EP009892.
XX PF
XX 28-AUG-2001; 2001WO-EP009892.
XX XX
XX (FARB ) BAYER AG.
XX PA
XX Tietjen K, Weidner M;
XX PI WPT; 2002-269010/31.
XX DR
XX Identifying plant target proteins for herbicidally active compounds,
XX PT comprising aligning and comparing nucleic acid or amino acid sequences
XX PT from plant with nucleic acid or amino acid sequences from non-plant
XX PT organisms.
XX XX
XX Claim 5; SEQ ID NO 2497; 26lpp + Sequence Listing; English.
XX PS
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
XX CC for herbicidally active compounds, comprising aligning and comparing
XX CC nucleic acid or amino acid sequences from plant with nucleic acid or
XX CC amino acid sequences from non-plant organisms using suitable search
XX CC parameters, where plant sequences having an E-value greater by a factor
XX CC of 3 than the E-value of most similar non-plant sequences are selected.
XX CC The polypeptides or nucleic acids encoding them are useful for
XX CC identifying modulators. The identified modulators are useful as
XX CC herbicides
XX XX
XX Sequence 359 AA;
XX SQ
Query Match 54.6%; Score 959; DB 5; Length 359;
Best Local Similarity 55.1%; Pred No. 3.6e-89;
Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;
QY 1 ETGAIDVRKVLXGVCHSDHMAKNDWGTSTYPIVPGHELVGVTVGVCKVKFKSWRQ 60
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
30 ETGEKDVRFKVLFCGICHSDLHMVKNWGMSTYPLVPGHELVGVTVGVCKVKFKTGEK 89
QY 61 -----GRCLHGRLEPTCNCIHLENYCPNLIQTYGSKYDGTWYGGYSNNMWTDE 113
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
90 VGVGCLVSSC-----GSCDSCTEGNEYCPKSIQTYGFFYDNTITYGGYSDHMYCEE 142
QY 114 HFTVRIPDNLPLDGAAPLCCAGITTYSPWRYYGLDKPGMHLGVWSPRRFRSRPPLNPLGL 173
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
143 GFVIRIPDNLPLDAAAPLCCAGITTYSPMKYHGLDKPGMHGV-----VGL 188
QY 174 WGRSLQSLVPLIKEGSGYGTSPALMHS-----LLRTDDQDQMEAAAMST 216
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
189 GG-----LGHGVKFAKAMGKVTIVISTSEKKRDEAINRLGADAFVSRDPKQIKDAMGT 243
QY 217 MDGIIDTVPAVRPLEPLISLLKTKGVTVGVIAVQPIDLVPFPLIIGKQVAGSAGSMK 276
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
244 MDGIIDTVSATHSLLPLLLGLLKHGKLVGVWGAPEKPLEPVMPLIFERKVMYGMISGSIK 303
QY 277 ETQMIDFAAEHNHTADIEVPIIDYLVNTAMERVVKQVRPFVIDVENTIL 326
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
304 ETQMIDMAGKNHTADIELISADYVNTAMERLEKADVRVRFVIDVANTIL 353
XX AC
XX 18-OCT-2000 (first entry)
XX DT
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 49819.
XX KW
XX Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS
XX Arabidopsis thaliana.
XX FN EP1033405-A2.
XX PD 06-SEP-2000.
XX XX
XX 25-FEB-2000; 2000EP-00301439.
XX PF
XX 25-FEB-1999; 99US-0121825P.
XX XX 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
XX PR 23-MAR-1999; 99US-0125788P.
XX PR 25-MAR-1999; 99US-0126264P.
XX PR 29-MAR-1999; 99US-0126785P.
XX PR 01-APR-1999; 99US-0127462P.
XX PR 06-APR-1999; 99US-0128234P.
XX PR 08-APR-1999; 99US-0128714P.
XX PR 16-APR-1999; 99US-0129845P.
XX PR 19-APR-1999; 99US-0130077P.
XX PR 21-APR-1999; 99US-0130449P.
XX PR 23-APR-1999; 99US-0130510P.
XX PR 28-APR-1999; 99US-0130891P.
XX PR 30-APR-1999; 99US-0131449P.
XX PR 30-APR-1999; 99US-0132048P.
XX PR 04-MAY-1999; 99US-0132407P.
XX PR 04-MAY-1999; 99US-0132484P.
XX PR 05-MAY-1999; 99US-0132485P.
XX PR 06-MAY-1999; 99US-0132486P.
XX PR 06-MAY-1999; 99US-0132487P.
XX PR 07-MAY-1999; 99US-0132863P.
XX PR 11-MAY-1999; 99US-0134256P.
XX PR 14-MAY-1999; 99US-0134218P.
XX PR 14-MAY-1999; 99US-0134219P.
XX PR 14-MAY-1999; 99US-0134221P.
XX PR 14-MAY-1999; 99US-0134370P.
XX PR 18-MAY-1999; 99US-0134768P.
XX PR 19-MAY-1999; 99US-0134941P.
XX PR 20-MAY-1999; 99US-0135124P.
XX PR 21-MAY-1999; 99US-0135353P.
XX PR 24-MAY-1999; 99US-0135629P.
XX PR 25-MAY-1999; 99US-0136021P.
XX PR 27-MAY-1999; 99US-0136392P.
XX PR 28-MAY-1999; 99US-0136782P.
XX PR 01-JUN-1999; 99US-0137222P.
XX PR 03-JUN-1999; 99US-0137528P.
XX PR 04-JUN-1999; 99US-0137502P.
XX PR 07-JUN-1999; 99US-0137724P.
XX PR 08-JUN-1999; 99US-0138094P.
XX PR 10-JUN-1999; 99US-0138540P.
XX PR 10-JUN-1999; 99US-0138847P.
XX PR 14-JUN-1999; 99US-0139119P.
XX PR 16-JUN-1999; 99US-0139452P.
XX PR 16-JUN-1999; 99US-0139453P.
XX PR 17-JUN-1999; 99US-0139432P.
XX PR 18-JUN-1999; 99US-0139454P.
XX PR 18-JUN-1999; 99US-0139455P.
XX PR 18-JUN-1999; 99US-0139456P.
XX PR 18-JUN-1999; 99US-0139457P.
XX PR 18-JUN-1999; 99US-0139458P.
XX PR 18-JUN-1999; 99US-0139459P.
XX PR 18-JUN-1999; 99US-0139460P.
XX PR 18-JUN-1999; 99US-0139461P.
```

```
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142902P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148313P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149336P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.

PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156594P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 50.1%; Score 881; DB 3; Length 360;
Best Local Similarity 49.1%; Pred. No. 3.6e-81;
Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRFKVLYCGVCHSDTHMAKNDWGTSTYPIVPGHELVGVVTEYGVCKYKESMRQ 60
Db 34 DNGENDVTVKILFCGVCHDTHHTKNDWGYSTYPPVPGHEIVGIAIKGVKNVTKFQGR 93
QY 61 -----GRCWLHGRLEPTCENCIHLHENYCPNLITQTYGSKYYDGTMTYGGYNNMWDE 113
Db 94 VGVGVISGSC-----QSCESCQDLENYCPQMSFTYNAIGSDCTKNYGGYSENVVDQ 146
QY 114 HPIVRIEDNLELDGHAPELLCAGITTYSPWRYGLDKPGMHLGSEWPRFRSRPPNLPGL 173
Db 147 RFVLRFPENLPSDGAPELLCAGITTYSPMKYTGTEAGKHLGV-----AGL 192
QY 174 WGSRLQSIAPPLIKEGSGYGTSPALMHS-----LLETDDOMEAAMST 216
Db 193 CG-----LGHVAVKIGKAFGLKVTVISSSTKAERINHLGADSEFLVTTDPQMKRAIGT 247
```

QY 217 MDGIITDTPAVRPLEPLISLLKTKNGKVTVGVIAVOPDLDPVPLIIGRWVAGSAIGMK 276
 Db 248 MDYIIDTISAVHALYPLGLKLVNGKLTALGPEKPLPELPMFPLVIGRWVAGSDVGGMK 307

QY 277 ETQEMIDFAAENHNTADIEVIPIDYINTAMERVKVVDVFRFRVIDVENTL 326
 Db 308 ETQEMLDFCAKHNITADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 357

RESULT 13
 ABB93303 ID ABB93303 standard; protein; 360 AA.
 AC ABB93303;
 XX
 DT 31-MAY-2002 (first entry)
 DE Herbicidally active polypeptide SEQ ID NO 2514.
 DE Herbicidal; plant; agriculture; herbicide.
 KW Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO200210210-A2.
 XX 07-FEB-2002.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX 28-AUG-2001; 2001WO-EP009892.
 XX (FARB) BAYER AG.
 XX Tietjen K, Weidner M;
 XX WPI; 2002-269010/31.
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.
 XX Claim 5; SEQ ID NO 2514; 261pp + Sequence Listing; English.
 XX The invention relates to identifying target proteins (ABB90790-ABB94016;
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides
 XX Sequence 360 AA;
 SQ

Query Match 50.1%; Score 881; DB 5; Length 360;
 Best Local Similarity 49.1%; Pred. No. 3.6e-81;
 Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRKVLKVCVCHSDHMAKNDWCTSTPTVPGHELVGVTVENGCKVKFKSWRO 60
 Db 34 DNGENDVTVKILFCVCHVDLHTIKNDGYSYVPVPGHEIVGIAIKVGNVTKFKEGDR 93

QY 61 -----GRCWLHGRLEPCTNCIHLHENYCNPLIQTYGSKYDGTMTYGGYSNNMVTDE 113
 Db 94 VGVGVISGSC-----QSCSCDQDLNFCYQMSFTYNAIGSDGTYKNGYSENIIVDQ 146

QY 114 HFIVRIPNLPLDGNAPLILCAGITITYSPWRYGLDKPGMHLGVFWPRFRSPPLNLPLGL 173
 Db 147 RFLRFPENLPSDSGAPLLCAGITITYSPMKYGYMTAGKHLGV-----AGL 192

QY 174 WGSRLQSLVPPLIKKGSGVGTSPALMHS-----LLRTDQDQWEAMST 216
 Db 193 CG-----LGHVAVKLGKAGLVTVVSSSTKAEAINHLGADSELVTTDPQMKAAIGT 247

QY 217 MDGIITDTPAVRPLEPLISLLKTKNGKVTVGVIAVOPDLDPVPLIIGRWVAGSAIGMK 276
 Db 248 MDYIIDTISAVHALYPLGLKLVNGKLTALGPEKPLPELPMFPLVIGRWVAGSDVGGMK 307

QY 277 ETQEMIDFAAENHNTADIEVIPIDYINTAMERVKVVDVFRFRVIDVENTL 326
 Db 308 ETQEMLDFCAKHNITADIELIKMDEINTAMERLAKSDVRYRVIDVANSL 357

RESULT 14
 AAG25506 ID AAG25506 standard; protein; 360 AA.
 AC AAG25506;
 XX
 DT 17-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 29600.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126783P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0128645P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132487P.
 PR 11-MAY-1999; 99US-0132863P.
 PR 14-MAY-1999; 99US-0134258P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 18-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142190P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147915P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151530P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 49.8%; Score 876; DB 3; Length 360;
Best Local Similarity 48.9%; Pred. No. 1.2e-80;
Matches 171; Conservative 52; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRFKVLYCGVCHSDJHMAKNDGCTSTYPIVPGHNLGVTEVCCKVKKKSWRQ 60
Db 34 DNGENDVTVKILFCGVCHTDLHTIKNDWGYSYYPVPGHEIYVGIATKVGKNVTKFKGDR 93
QY 61 -----GRCWLHGRLRPTCENCIHLNENYCPNLIQTYGSKYDGTMTVGGYSNNWVTD 113
Db 94 VCVGVISGSC-----QSCSCDQDLENYCPQMFTYNAIGSDGTYGKYGSENIIVDQ 146
QY 114 HFIVRIPDNLEFDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRPRPPLNLPGL 173
Db 147 RFVLFPPNLSGAPLLCAGITTYSPWRYGLDKPGMHLGVWEPFRPRPPLNLPGL 173
QY 174 WGRLOSLVPPVPLIKGSGYGTSPALMHS-----LLRTDQDMEANMT 216
Db 193 GG-----LGHVAVKIGKAPGLKVTVISSSTKAEAINHLGADSLFTVTDPOKMGAAIGT 247
QY 217 MDGIIDTVPVPRPLFLSLKTKNGKWTVTGIAVOPDLPVPLIIGRKMVAGSAGCK 276
Db 248 MDYIIDTISAVHALYPLGLKLVNGKLIAGLPEKPELPEPVLGRKMGVGGSDVGGCK 307
QY 277 ETOEMIDFAENITADIEVTPIDVLTAMERVKVKKDVRFRVIDVENTL 326
Db 308 ETOEMLDFCAKHNITADIELIKMDRINTAMERLAKSDVRYRFINVANS 357

RESULT 15
AAG29471
ID AAG29471 standard; protein; 306 AA.
AC AAG29471;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 35072.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
XX
PO 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-00301439.
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134220P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134222P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135622P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140695P.
PR 29-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143542P.
PR 15-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144633P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.

```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0147038P.
PR 03-AUG-1999; 99US-0147038P.
PR 03-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 49.5%; Score 871; DB 3; Length 306;
Best Local Similarity 54.7%; Pred. No. 3e-80;
Matches 176; Conservative 43; Mismatches 65; Indels 38; Gaps 5;

QY 23 MAKNDWGTSTYPIVPGHELVGVVTEVGCKVKPKSM-ROGRCLHGRLRPTCNCIHHLE 81
DB 1 MAKNEWGLTTPVPGHEIVGVVTEVGAKVKFNAGDKVGVGWSGCR-SCDSCNDGDE 59
QY 82 NYCPLNIQYGSKYDGYNTYGSYSNNMVTDEHFIIVRIIPNLELDGAAPLLCAGITTYSP 141
DB 60 NYCPRMILTSGAKNFDDTWHGYSYDHMVCAEDFIIRIPNLELDGAAPLLCAGITTYSP 119
QY 142 WRYYGLDKQGMHLGVWEPFRFRSRPPNLPGLWGSRLQSLVPEPLIXEGGSYGTSPALMHS 201
DB 120 MKYHGLDKPQMHIGV-----VGLGG-----LGHVAVKFAKAMGTKVTVIST 160
QY 202 -----LLRTDQDOMEAAMSTMDGIIDVPAVRPLEPLISLKTNGKV 244
DB 161 SERKDEAVTRLGADAFVSRDEKQMKDAMGTMDGIIDIVSATHPLPLGLLKNKGLV 220
QY 245 TVGIAVQPLDLVPFPLIIGRKVAGSAIGMKETQEMIDFAAEHNTADIEVIPIDYANT 304
DB 221 MVGAPAEPLPPLVPLIFGRKVVGVGSGIKETQEMVDLAGKHNTADIELISADYVNT 280
QY 305 AMERVVKDVRFRFVIDVENTL 326
DB 281 AMERLAKADVRYRFVIDVANTM 302

Search completed: July 6, 2004, 13:36:45
Job time : 44.3213 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:34:55 ; Search time 13.8278 Seconds
(without alignments)
1217.121 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758
Sequence: 1 ETGATDVRFKVLICGVCHSD.....ERVVKDVRFRFVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1007	57.3	355	4	US-09-501-115-12
2	869	49.4	360	4	US-09-501-115-4
3	832	47.3	371	4	US-09-501-115-14
4	800	45.5	361	4	US-09-501-115-8
5	790	44.9	366	4	US-09-501-115-10
6	732.5	41.7	364	4	US-09-501-115-2
7	717	40.8	358	4	US-09-501-115-6
8	665	37.8	360	4	US-09-501-115-32
9	648	36.9	358	4	US-09-501-115-30
10	630.5	35.9	353	4	US-09-328-352-7473
11	579.5	33.0	410	4	US-09-252-991A-25647
12	578	32.9	357	4	US-09-501-115-20
13	567.5	32.3	358	4	US-09-489-039A-10508
14	559	31.8	292	4	US-09-501-115-28
15	379	21.6	360	4	US-09-501-115-34
16	366.5	20.8	374	4	US-09-433-039A-8925
17	356	20.3	358	4	US-09-328-352-4988
18	316	18.0	344	4	US-09-134-001C-5343
19	310	17.6	349	2	US-07-857-224B-95
20	303	17.2	186	4	US-09-501-115-16
21	299.5	17.0	150	4	US-09-501-115-24
22	298.5	17.0	172	4	US-09-501-115-46
23	297.5	16.9	465	4	US-09-252-991A-18919
24	296	16.8	155	4	US-09-501-115-48
25	273.5	15.6	375	4	US-09-134-000C-5937
26	270	15.4	206	4	US-09-540-236-3768
27	265	15.1	346	4	US-09-489-039A-9885

ALIGNMENTS

RESULT 1
US-09-501-115-12
; Sequence 12, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BBL328 US NA
; CURRENT APPLICATION NUMBER: US/09/501.115
; EARLIER FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1993-February-10
; NUMBER OF SEQ ID NOS: 48
; SEQ ID NO 12
; SOFTWARE: Microsoft Office 97
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Glycine max
US-09-501-115-12

Query Match 57.3%; Score 1007; DB 4; Length 355;
Best Local Similarity 59.6%; Pred. No. 1.2e-100;
Matches 201; Conservative 39; Mismatches 73; Indels 24; Gaps 5;
Qy 1 ETGATDVRFKVLICGVCHSDIHMANKNDWGTSTYPIVPGHELGVGVTVGCKVKPK-SNR 59
Db 28 ETSEKDLVFKVQYCGICHSIDHMLKNEWNTYPLVFGHEIAGVTVTEVGSVKQKFGVDR 87
Qy 60 QGRCLMGRRLRPTCNCIHLNVCNLIQYGYKYDGTMTYGYSNMVTDEHFVRI 119
Db 88 VGVCMIGSCR-SCSCDLENLYCPKMLTYGVKFDGTHGYSIDMVADEHFVRI 146
Qy 120 PDNLPLDGAAPLLCAGITTYSPWRYGLDKFGEHMLGVWEPFRSRPPLNPLGLWGSRLQ 179
Db 147 PDNLPLDGAAPLLCAGITTYSPWRYGLDKFGEHMLGVWEPFRSRPPLNPLGLWGSRLQ 194
Qy 180 SLVPLPKEGSGVGTSPALMHSLLRT-----DQDQEAAMSTWDGIIDTVPVAVRP 229
Db 195 KFAKALGANVTIVISTSPNKKKEATENTGADSFVVSREQDQQAQVMTGDIIDTIVSAVHP 254
Qy 230 LEPLISLTKNGKVVTVGVIAVQPLDLPVFPILLIGSKMVGSAIGCMKETQEMIDFAEHN 289
Db 255 LVPLIGLKHGHKLVMVGAPEKPLELFPVSLLMGRKMWGSSIGCMKETQEMIDFAKHG 314
Qy 290 ITADIEVPIIDYLNLTAMERVKKDVRFRFVIDVENTL 326
Db 315 VKPDIEVPIIDYLNLTALERLAKADVKRFRVIDIGNTL 351

RESULT 2
 US-09-501-115-4
 ; Sequence 4, Application US/09501115
 ; Patent No. 6552249
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Fader, Gary M.
 ; APPLICANT: Rafalski, Antoni
 ; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
 ; FILE REFERENCE: BB1328 US NA
 ; CURRENT APPLICATION NUMBER: US/09/501,115
 ; CURRENT FILING DATE: 2000-02-09
 ; EARLIER APPLICATION NUMBER: 60/119,585
 ; EARLIER FILING DATE: 1999-February-10
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 4
 ; LENGTH: 360
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-09-501-115-4

Query Match 49.4%; Score 869; DB 4; Length 360;
 Best Local Similarity 51.3%; Pred. No. 1.1e-85;
 Matches 174; Conservative 47; Mismatches 88; Indels 30; Gaps 4;
 QY 1 ETGATDVRPKVLYCGVCHSDIHMKNKDWGTYPIVPGHVLGVVTEVCGCKVKKSWRQ 60
 DB 34 ENGVDVDTLKILFCGVCHSDHLTKNDKNGFTTYPVPGHVLGVVTKVGNVKNPKVGDK 93
 QY 61 GRCLMHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDDEHFIVRP 120
 DB 94 VGVGVIVSEKCECQQDLESYCYCPVPTYNYPYDGTGRTKRGYSNIMVHVQRYLRRP 153
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 180
 DB 154 ENLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 197
 QY 181 LVPPPLKEGG---SVGTSALMHS-----LLRTDQDQMEAMSTMDGIIDTTPA 226
 DB 198 VAIKAKAFGLKLLSSSPENKQAEIRDLGADSLVSSDPAMKVALGTMDYIIDTISA 257
 QY 227 VRPLPLELSLTKNGKVVTVGIAVQPLDLPVPLIIGRMVAGSSAIGGMKTKQEMIDFAA 286
 DB 258 VHSLLPLGLLKLNGKLVTVGLPNKPLEPIPLVAGRKLIGSNFNGGIGKEQEMIDFCA 317
 QY 287 EHNITADIEVIPIDYINTAMERVVKDVRFRFVIDVENT 325
 DB 318 KHNITADIELKMDQINTAMERLSKADVKYRFVIDVANS 356

RESULT 3
 US-09-501-115-14
 ; Sequence 14, Application US/09501115
 ; Patent No. 6552249
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Fader, Gary M.
 ; APPLICANT: Rafalski, Antoni
 ; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
 ; FILE REFERENCE: BB1328 US NA
 ; CURRENT APPLICATION NUMBER: US/09/501,115
 ; CURRENT FILING DATE: 2000-02-09
 ; EARLIER APPLICATION NUMBER: 60/119,585
 ; EARLIER FILING DATE: 1999-February-10
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 14
 ; LENGTH: 371
 ; TYPE: PRT
 ; ORGANISM: Triticum aestivum
 US-09-501-115-14

Query Match 47.3%; Score 832; DB 4; Length 371;
 Best Local Similarity 48.8%; Pred. No. 1.2e-81;
 Matches 168; Conservative 52; Mismatches 84; Indels 40; Gaps 5;
 QY 2 TGATDVRPKVLYCGVCHSDIHMKNKDWGTYPIVPGHVLGVVTEVCGCKVKKSW-RQ 60
 DB 40 TGDDDDVIXILYCGICHSDLSIKNDKNAKYPMPGHEIAGEVTEVGVNVTKFKAGDRV 99
 QY 61 G-RCLMHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDDEHFIVR 119
 DB 100 GVGCMVNS--CQSCSCDKGPHNCPGMIFTYNSVDRDGTGRTTHGGYSNNVWVHERFWVF 157
 QY 120 PDLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGLWGSRLQ 179
 DB 158 PDAMPDLKGAFLLCAGITTYSPMKYHGLNAPGMHLGV-----LGLGLG----- 201
 QY 180 SLVPPPLKEGSGYTSALMHS-----LLRTDQDQMEAMSTMDGIID 222
 DB 202 --HVAVKFGKAFGMKVTVISSSPGKQKQEALERLGADAFVVSADAKATMTMDGIIN 258
 QY 223 TVPVRPLEPLISLTKNGKVVTVGIAVQPLDLPVPLIIGRMVAGSSAIGGMKTKQEMI 282
 DB 259 TVSANVPMAPLFGLLKNPKMIMVGLPEKPIEVPPFALVARNKTLGSCIGMRDQEWL 318
 QY 283 DFAAEHNITADIEVIPIDYINTAMERVVKDVRFRFVIDVENTL 326
 DB 319 DLAAKHGVTADIEVGAENVNTAMERLAKADVYRFVIDIANL 362

RESULT 4
 US-09-501-115-8
 ; Sequence 8, Application US/09501115
 ; Patent No. 6552249
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Fader, Gary M.
 ; APPLICANT: Rafalski, Antoni
 ; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
 ; FILE REFERENCE: BB1328 US NA
 ; CURRENT APPLICATION NUMBER: US/09/501,115
 ; CURRENT FILING DATE: 2000-02-09
 ; EARLIER APPLICATION NUMBER: 60/119,585
 ; EARLIER FILING DATE: 1999-February-10
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 8
 ; LENGTH: 361
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (39)
 ; FEATURE:
 ; NAME/KEY: UNSURE
 ; LOCATION: (78)
 US-09-501-115-8

Query Match 45.5%; Score 800; DB 4; Length 361;
 Best Local Similarity 47.7%; Pred. No. 3.5e-78;
 Matches 165; Conservative 52; Mismatches 83; Indels 46; Gaps 7;
 QY 2 TGATDVRPKVLYCGVCHSDIHMKNKDWGTYPIVPGHVLGVVTEVCGCKVKKSW-RQ 60
 DB 33 TGDDDDVIXILYCGICHSDLSIKNDKNAKYPMPGHEIAGEVTEVGVNVTKFKAGDRV 92
 QY 61 G-RCLMHGLRPTCNCIHLHENYCPNLQTYGSKYDGTMTYGGYSNNMTDDEHFIVR 119
 DB 93 GVGCMVNS--CQSCGSCDEDFENSCRAVIFTYNSVDRDGTGTYGGYSSVWVHERFWVF 150
 QY 120 PDLPLDGAAPLLCAGITTYSPWRYGLDKPGMHGVWPRFRSRPPLMLPGL----- 173
 DB 151 PDAMPDLKGAFLLCAGITTYSPMKYHGLNVPKGV-----MGLGLGHVAYK 199

```

; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BBI328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-501-115-2

Query Match      41.7%; Score 732.5; DB 4; Length 364;
Best Local Similarity 43.8%; Pred. No. 7.6e-71;
Matches 152; Conservative 55; Mismatches 81; Indels 59; Gaps 8;

QY 174 -----WGRISQSLVPLIKEGSGVTSPALMHSLL-----RTDQDQMEAAAMSTMDDG 219
DB 200 FAKAFGAKVTVI-----STSPGKQKQZALDRIGADAFVVSQDAEEMKAAASTMDG 248
QY 220 IITDTPAVRPLEPLISLLKTKNGVVTGVIAGVQPLDLPVPEPLIIGRKWVAGSAIGMKETQ 279
DB 249 IINTVSANUTLAPYMGLLKPKNGKIMVGLPTKLEIPPPDLIIGNKTKLAGSCIGGMRDQ 308
QY 280 EMIDFAAEHNITADIEVIPIDYLTAMERVVKDVRFRFVIDVENT 325
DB 309 EMINVAAGHGTADIELVAADYVNTAMERLAKADVRFRFVIDIGNT 354

RESULT 5
US-09-501-115-10
; Sequence 10, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BBI328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-501-115-10

```

```

Query Match      44.9%; Score 790; DB 4; Length 366;
Best Local Similarity 46.3%; Pred. No. 4.4e-77;
Matches 161; Conservative 56; Mismatches 88; Indels 38; Gaps 5;

QY 2 TGATDVRFKVLYCGVCHSDIHMANKNDWGTSTYPIVPGHELGVVTVGCKVKFKSW-RQ 60
DB 38 TGDDDDVAIKLFCGICHSDLHCINKWKSIIYPLVPGHELGVVTVGCKVKFKAGDRV 97
QY 61 GRCLWHLGRRLRPTCNCIHLHLENYCPNLIQTYGKYVDGTYGYSNNMVTDEHFVIRP 120
DB 98 GVCWMVNSCR-SCSCNNGFENHCPESVFTYNSVDKDGTVTYGYSNNMVTDEHFVIRP 156
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRSRPLNLPGLWGSRLQS 180
DB 157 EAPLWDGAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRSRPLNLPGLWGSRLQS 199
QY 181 LVPPLIKEGSGVTSPALMHS-----LLRTDQDQMEAAAMSTMDDGIIDT 223
DB 200 --HVAKZARAFGLKVTYVTSPPGKKEALERLIGADAFVVSSEAEKAAASTMDGVINT 257
QY 224 VPVRLPLEPLISLLKTKNGVVTGVIAGVQPLDLPVPEPLIIGRKWVAGSAIGMKETQ 283
DB 258 VSAINTMAPYLLALKPKNGKIMVGLPTKLEIPPPDLIIGNKTKLAGSCIGGMRDQ 317
QY 284 FAEHNITADIEVIPIDYLTAMERVVKDVRFRFVIDVENTL 326
DB 318 LAAGHGVTDIEVGADVDVNTAMERLAKADVRFRFVIDVGNLT 360

```

```

RESULT 6
US-09-501-115-2
; Sequence 2, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BBI328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; CURRENT FILING DATE: 2000-02-09
; EARLIER APPLICATION NUMBER: 60/119,585
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-501-115-6

Query Match      40.8%; Score 717; DB 4; Length 358;
Best Local Similarity 43.1%; Pred. No. 3.6e-69;
Matches 147; Conservative 61; Mismatches 93; Indels 40; Gaps 6;

QY 6 DVRFKLYCGVCHSDIHMANKNDWGTSTYPIVPGHELGVVTVGCKVKFKK-----SW 58
DB 33 DVTIKVLCGLCHTDLHVIKNKFNKAKYPPVPGHEIVGVVTVGSGVTSFKPGDVTGCVGY 92

```

```

QY 59 RQRCWLHGRRLPTECNCIHLNCPNLQIOTYGSKYD---GTMTYGGYNNMVTDEH 114
DB 93 FVDS-----RSCDSCKSYESCQPLVETNSVSLDDGCGATTGGFSDALVHOR 145
QY 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL- 173
DB 146 YVVRVPASLPAPGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL- 194
QY 174 -----WGRSLQSLVPLIKGSGYSTSPALMHSLLRTDQDMEAAAMSTMDGIIDTV 224
DB 195 HLAVERFGKAPGMKVTVISTSLGRDEALGRGADAP-LVSRDPEQRAAAGTLGVIDTV 253
QY 225 PAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLIIGRKMVAGSAIGMKETQRMIDF 284
DB 254 SADHPVPLDLLKPMQGVVVVGLTKPLQVPAFSLVAGGKRVAGSAGSGVCECQMLDF 313
QY 285 AAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDVENT 325
DB 314 AGEHGITADVEVGMVDYNTATQRLERNDRVRYFVVDVASS 354

```

RESULT 8

US-09-501-115-32

```

; Sequence 32, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: B31328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; EARLIER FILING DATE: 2000-02-09
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-501-115-32

```

```

Query Match 37.8%; Score 665; DB 4; Length 360;
Best Local Similarity 40.6%; Pred. No. 1.6e-63;
Matches 141; Conservative 57; Mismatches 97; Indels 52; Gaps 7;

QY 1 ETGATDVRPKVLYCGVCHSDIHMANKDWTSTPIVPGHVLGVTVGVCKVKFKSWR- 59
DB 32 KTGPDVVLKVKYCGICHTDHVQKNDLGASKYPMVPGHEVGVGVGVGVGVGVGV 91
QY 60 -----QGRCLHGRRLPTECNCIHLNCPNLQIOTYGSKYDGTMTYGGYNNMVTDE 113
DB 92 VGVGVVGC-----RDCRCKANVQYCNKKLWSNDVTDKPTQGGFASAMVVDQ 144
QY 114 HFVIRIPDNLPLDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL 173
DB 145 KPVVKIPAGLAPEQAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL 193
QY 174 WGRSLQSLVPLIKGSGYSTSPALMHS-----LLRTDQDMEAAAMST 216
DB 194 GHMG-----VKYAKSMGHVTVISSNKKRAEMDDLGADAYLVSSDTPDQMAAADS 245
QY 217 MDGIIDTVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLI-IGRKMVAGSAIGM 275
DB 246 LDYIIDTVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLI-IGRKMVAGSAIGM 304
QY 276 KETQRMIDFAAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDV 322
DB 305 DETEVLQFCVDKGLTQSEVVMVDYNTATQRLERNDRVRYFVVDV 351

```

RESULT 9

```

Query Match 35.9%; Score 630.5; DB 4; Length 353;
Best Local Similarity 39.9%; Pred. No. 8.6e-60;

```

```

US-09-501-115-30
; Sequence 30, Application US/09501115
; Patent No. 6552249
; GENERAL INFORMATION:
; APPLICANT: Caboon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: B31328 US NA
; CURRENT APPLICATION NUMBER: US/09/501,115
; EARLIER FILING DATE: 2000-02-09
; EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 30
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (201)
US-09-501-115-30

```

```

Query Match 36.9%; Score 648; DB 4; Length 358;
Best Local Similarity 41.0%; Pred. No. 1.1e-61;
Matches 141; Conservative 58; Mismatches 103; Indels 42; Gaps 9;

QY 2 TGATDVRPKVLYCGVCHSDIHMANKDWTSTPIVPGHVLGVTVGVCKVKFKSWRQ 61
DB 32 TGPDDVYIKVHYCGICHSDIHMANKDWTSTPIVPGHVLGVTVGVCKVKFKSWRQ 88
QY 62 RCWLHGRRLPTECNCIHLNCPNLQIOTYGSKYDGTMTYGGYNNMVTDEHFI 118
DB 89 ELVGVGLLVGCRNCPQDQDIENYCSKKIWSNDVYDVKETQGFRTMTVEQKFWVK 148
QY 119 IPDNLPLDGAAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL- 176
DB 149 IPEGLAPEQVAPLLCAGITTYSPWRYXGLDKPGMHLGVEMPRFRSRPLNPLGL- 197
QY 177 RQSLVPLIKGSGYST-----SPALMH-----SLRTDQDMEAAAMSTMDGIID 222
DB 198 KXAL-----GHVTVISSDKKKQEALEHLGADQYLVSSDATAMQEAADSLDIID 250
QY 223 TPVAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVFPPLI-IGRKMVAGSAIGMKETQ 281
DB 251 TPVCHPPLPPLISLKTNGKVVTVGIAVQPLDLPVFPPLI-IGRKMVAGSAIGMKETQ 309
QY 282 IDFAAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDVENT 325
DB 310 LEFMKEKGLSSMIEVMNDYINKAFERLEKNDVRVRYFVVDVVKGS 353

```

RESULT 10

```

US-09-328-352-7473
; Sequence 7473, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7473
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7473

```



```
Query Match      31.8%; Score 559; DB 4; Length 292;
Best Local Similarity 41.6%; Pred. No. 3.7e-52;
Matches 124; Conservative 44; Mismatches 88; Indels 42; Gaps 9;
```

QY	1	ETGATDVRKVIYCVCYCHSDIHMAKDNGTSTPIYPVGHVLGVVTVEGCKVKFKSW-R	59
Db	16	ENGVDWTIKILCYGICHTDLHAKNKGWITMYPVPVPGHEIIGVTKGRDVKNVFKEGDR	75
QY	60	QG--RCLWHLGRLRPTCENCILHHLENYCNPLIQTYGSKYYDGITMTYGGYSNMWTDHFIVR	118
Db	76	VGVGCLLAASCLE--CHECKTDQENYCEKQFYVINGVFDGSITYGGYSQIFVDADYRYVTH	133
QY	119	IDPNIPLDGAAPLLCAGITITYSPRWYGL-DKPGMHLGVMFRRPRSRPLNDPLGLWGSR	177
Db	134	IPENTAMDAAAPLACAGITVFNFLPKDHVLVASPKKXIGV-----VGJGG--	177

Search completed: July 6, 2004, 13:40:16
Job time : 14.8278 secs

QY 1 ETGATDVREKVLVCGVCHSDIHMAKNDWGTSTYPIVPGHELGVVTEVGC

Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFMSYPLVPGHIEVGEVTEVSGVKVKNVGDVK 94
Qy 61 -----GRCWLHGRRLPTCENCIEHLENCPMLIOTYKSYDGTMTYGGYSNNMTDEH 114
Db 95 GVGCLVGAC-----HSCESCANDLENYCPKMLITYASIHDTGTTTGGYSDDHVMANER 147
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLGVWEPFRFRPLNLPGLW 174
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGLDEPKHIGI-----VGLG 193
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220
Db 194 G--LGHVAVKFAKFGSKVTYISTSPSKKEALKNFGADSFVSRDQEQMRAAGTLDDGI 251
Qy 221 IDTVPAVRPLEPLISLLKTNKGVVTVGIAVQPLDLPVFPPLIIGRKWVAGSAIGGMKETQE 280
Db 252 IDTVSAVHPLPLFGLKSHGKLLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGGMKETQE 311
Qy 281 MIDFAAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDVENTL 326
Db 312 MIDFAAKHNITADIEVISTDYINTAERLAKNDVRYRFRFVIDVGNL 357

RESULT 5

US-09-947-027-2
; Sequence 2, Application US/09947027
; Patent No. US20020124281A1
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent Lee C.
; APPLICANT: Li, Laigeng
; TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION
; FILE REFERENCE: 066040-9718
; CURRENT APPLICATION NUMBER: US/09/947,027
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: 60/230,086
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 362
; TYPE: PRT
; ORGANISM: aspen populus tremuloides
US-09-947-027-2

Query Match 55.7%; Score 979; DB 9; Length 362;
Best Local Similarity 56.6%; Pred. No. 7.2e-94;
Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

Qy 2 TGATDVRFKVLYCGVCHSDIHMKNWGTSTYPIVPGHIEVGEVTEVSGVKVKNVGDVK 60
Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFMSYPLVPGHIEVGEVTEVSGVKVKNVGDVK 94
Qy 61 -----GRCWLHGRRLPTCENCIEHLENCPMLIOTYKSYDGTMTYGGYSNNMTDEH 114
Db 95 GVGCLVGAC-----HSCESCANDLENYCPKMLITYASIHDTGTTTGGYSDDHVMANER 147
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLGVWEPFRFRPLNLPGLW 174
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGLDEPKHIGI-----VGLG 193
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220
Db 194 G--LGHVAVKFAKFGSKVTYISTSPSKKEALKNFGADSFVSRDQEQMRAAGTLDDGI 251
Qy 221 IDTVPAVRPLEPLISLLKTNKGVVTVGIAVQPLDLPVFPPLIIGRKWVAGSAIGGMKETQE 280
Db 252 IDTVSAVHPLPLFGLKSHGKLLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGGMKETQE 311
Qy 281 MIDFAAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDVENTL 326
Db 312 MIDFAAKHNITADIEVISTDYINTAERLAKNDVRYRFRFVIDVGNL 357

RESULT 6

US-10-091-009-2
; Sequence 2, Application US/10091009
; Publication No. US20020138870A1
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent Lee C.
; APPLICANT: Li, Laigeng
; TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND
; TITLE OF INVENTION: COMPOSITION, AND
; FILE REFERENCE: 066040-9718
; CURRENT APPLICATION NUMBER: US/10/091,009
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/947,027
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: 60/230,086
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 362
; TYPE: PRT
; ORGANISM: aspen populus tremuloides
US-10-091-009-2

Query Match 55.7%; Score 979; DB 13; Length 362;
Best Local Similarity 56.6%; Pred. No. 7.2e-94;
Matches 196; Conservative 39; Mismatches 67; Indels 44; Gaps 6;

Qy 2 TGATDVRFKVLYCGVCHSDIHMKNWGTSTYPIVPGHIEVGEVTEVSGVKVKNVGDVK 60
Db 35 TGEEDVRFKVLYCGICHSDLSHDKNDWGFMSYPLVPGHIEVGEVTEVSGVKVKNVGDVK 94
Qy 61 -----GRCWLHGRRLPTCENCIEHLENCPMLIOTYKSYDGTMTYGGYSNNMTDEH 114
Db 95 GVGCLVGAC-----HSCESCANDLENYCPKMLITYASIHDTGTTTGGYSDDHVMANER 147
Qy 115 FIVRIPDNLPLDGAAPLLCAGITTYSPWRYGGLDKPGHGLGVWEPFRFRPLNLPGLW 174
Db 148 YIIRFPDNNPLDGGAPLLCAGITTYSPKLYFGLDEPKHIGI-----VGLG 193
Qy 175 GSRLQSLVPLIKEGGS-----YGTSPALMHSLLRT-----DQDQMEAMSTMDGI 220
Db 194 G--LGHVAVKFAKFGSKVTYISTSPSKKEALKNFGADSFVSRDQEQMRAAGTLDDGI 251
Qy 221 IDTVPAVRPLEPLISLLKTNKGVVTVGIAVQPLDLPVFPPLIIGRKWVAGSAIGGMKETQE 280
Db 252 IDTVSAVHPLPLFGLKSHGKLLILVGAPEKPLELPFAFSLIAGRKIVAGSGIGGMKETQE 311
Qy 281 MIDFAAENHITADIEVIPIDYINTAMERVVKDVRFRFVIDVENTL 326
Db 312 MIDFAAKHNITADIEVISTDYINTAERLAKNDVRYRFRFVIDVGNL 357

RESULT 7

US-10-424-599-266973
; Sequence 266973, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266973
; LENGTH: 361
; TYPE: PRT

; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 36617
 ; LENGTH: 365
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3040-049-D9_FLI.pep
 US-10-425-114-36617

Query Match 50.3%; Score 884; DB 12; Length 365;
 Best Local Similarity 50.6%; Pred. No. 7.1e-84;
 Matches 173; Conservative 49; Mismatches 84; Indels 36; Gaps 3;

 QY 1 ETGATDVRFKVLGCGVCHSDIHKAKNDWGTSYPIVPGHGLGVVTVGVCKKFKPSWRQ 60
 Db 39 ENGVDVDTLKILFCGVCHSDLTILKNDWGFTTYPVPGHEIVGVTVGVNNVKNFKVGD 98
 QY 61 GRCLWHLRLPTCENCIHLENYCPNLITQYSGKYDGTMTYGGYSNNMVTDEHFIVRIP 120
 Db 99 VGVGVIVESCKECCQODLENYCPREFTYNSPYDGTTRTQGGYSNIVVWHQYVLRFP 158
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRSRPPLNPLGLWGSRLQS 180
 Db 159 ENPLDAGAPLLCAGITTYSPMKYTGTEPGKHLGV-----AGLGG----- 199
 QY 181 LVPPLIKEGSGYGTSPALMHS-----LLRTDQDMEAAAMSTMDGIIDT 223
 Db 200 LGHVAIKLAKAFGLKVTWISSPNKQAEADRLGADFLVSSDPKMKAAALGTMDYIIDT 259
 QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKMVAGSAIGMKETQEMID 283
 Db 260 ISAVHSLIPLGLKNGKLVTVGLPNKPLELPFPLVAGRKLGGNFGGLKXETQEMLD 319
 QY 284 FAEHNITADIEVIPIDYLTAMERVVKXDVRRFRFVIDVENT 325
 Db 320 FCGKHNTADIELIKMQDQINTAMERLSRADVKYRFVIDVASS 361

RESULT 11
 US-10-425-114-54709
 ; Sequence 54709, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 54709
 ; LENGTH: 365
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: JC-GMFL:2220073B02_FLI.pep
 US-10-425-114-54709

Query Match 50.3%; Score 884; DB 12; Length 365;
 Best Local Similarity 50.6%; Pred. No. 7.1e-84;
 Matches 173; Conservative 49; Mismatches 84; Indels 36; Gaps 3;

 QY 1 ETGATDVRFKVLGCGVCHSDIHKAKNDWGTSYPIVPGHGLGVVTVGVCKKFKPSWRQ 60
 Db 39 ENGVDVDTLKILFCGVCHSDLTILKNDWGFTTYPVPGHEIVGVTVGVNNVKNFKVGD 98
 QY 61 GRCLWHLRLPTCENCIHLENYCPNLITQYSGKYDGTMTYGGYSNNMVTDEHFIVRIP 120
 Db 99 VGVGVIVESCKECCQODLENYCPREFTYNSPYDGTTRTQGGYSNIVVWHQYVLRFP 158
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRSRPPLNPLGLWGSRLQS 180
 Db 159 ENPLDAGAPLLCAGITTYSPMKYTGTEPGKHLGV-----AGLGG----- 199
 QY 181 LVPPLIKEGSGYGTSPALMHS-----LLRTDQDMEAAAMSTMDGIIDT 223
 Db 200 LGHVAIKLAKAFGLKVTWISSPNKQAEADRLGADFLVSSDPKMKAAALGTMDYIIDT 259
 QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKMVAGSAIGMKETQEMID 283
 Db 260 ISAVHSLIPLGLKNGKLVTVGLPNKPLELPFPLVAGRKLGGNFGGLKXETQEMLD 319
 QY 284 FAEHNITADIEVIPIDYLTAMERVVKXDVRRFRFVIDVENT 325
 Db 320 FCGKHNTADIELIKMQDQINTAMERLSRADVKYRFVIDVASS 361

RESULT 12
 US-10-424-559-152604
 ; Sequence 152604, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J
 ; APPLICANT: Kovalic, David K
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 152604
 ; LENGTH: 360
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_108826C.1.pep
 US-10-424-559-152604

Query Match 50.1%; Score 880; DB 12; Length 360;
 Best Local Similarity 50.6%; Pred. No. 1.8e-83;
 Matches 173; Conservative 49; Mismatches 84; Indels 36; Gaps 3;

 QY 1 ETGATDVRFKVLGCGVCHSDIHKAKNDWGTSYPIVPGHGLGVVTVGVCKKFKPSWRQ 60
 Db 34 ENGVDVDTLKILFCGVCHSDLTILKNDWGFTTYPVPGHEIVGVTVGVNNVKNFKVGD 93
 QY 61 GRCLWHLRLPTCENCIHLENYCPNLITQYSGKYDGTMTYGGYSNNMVTDEHFIVRIP 120
 Db 94 VGVGVIVESCKECCQODLENYCPREFTYNSPYDGTTRTKGGYSNIMVWHQYVLRFP 153
 QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVWPRFRSRPPLNPLGLWGSRLQS 180
 Db 154 ENPLDAGAPLLCAGITTYSPMKYTGTEPGKHLGV-----AGLGG----- 194
 QY 181 LVPPLIKEGSGYGTSPALMHS-----LLRTDQDMEAAAMSTMDGIIDT 223
 Db 195 LGHVAIKLAKAFGLKVTWISSPNKQAEADRLGADFLVSSDPKMKAAALGTMDYIIDT 254
 QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKMVAGSAIGMKETQEMID 283

[illegible]

```

RESULT 13
US-10-425-114-44445
; Sequence 44445, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44445
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700954121 FLI.pcp
US-10-425-114-44445

```

RESULT 14
US-10-357-886--4
; Sequence 4, Application US/10357886
; Publication No. US20030159170A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Fader, Gary M.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Cinnamyl-Alcohol Dehydrogenase Homologs
; FILE REFERENCE: BB1328 US NA
; CURRENT APPLICATION NUMBER: US/10/357.886

```

; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: US/09/501,115
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/119,585
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-February-10
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Glycine max
US-10-357-886--4

      Query Match          49.4%; Score 869; DB 14; Length 360;
      Best Local Similarity 51.3%; Pred. No. 2.6e-82;
      Matches 174; Conservative 47; Mismatches 88; Indels 30; Gaps 4;

Qy 1  ETGATDVRFKVLGVGCHSDIHMAKNDWGTSTYPIVPGHELVGVTVEGCKVKFKFSWQ 60
Db 34  ENGVDVDTLKILFCGVCHSDLHTLKNWDGFTYPPVVFEGHEIVGVVTKVGNKVQNFVGDK 93

Qy 61  GRCHWGLRLPTCNCIHLEHLCNLIQTSKYVGDGTWTCYGYSSNNMTDEHFTVIRP 120
Db 94  VGVGVIVSECKECSQDLSEYCPRPVFTYNSPYDGTTRKRGYSNIMVWHQRYVYRFP 153

Qy 121  DNLPLDGAPELLCAGITTYSPFMYCYGLDPEGMHLGVENPFRSPRPLKLPGLWGSRLQS 180
Db 154  ENIPLDAGAPLLCAGITTYSPMKYGYGTMTEFGKHLG-----AGLGG--LGH 197

Qy 181  LVPLPLKEGG-----SYGTSFALMHS-----LLRTDQDQMEAMSTMDGIDTVPA 226
Db 198  VAIKLAKAFGLKLLSSSSFNKQAEIDLGDGSFLVSSDPAMKVKVALGTMVDITISA 257

Qy 227  VRPEPLISLKTNGKVVTVYGIAPQDLDFVPLIIGROWVAGSAICGMKETQEMIDFAA 286
Db 258  VHSILPILGLKXNGKLVTVGLPNKPELELPFPLVAGRKLIGGSNFGGIKETQEMLDFA 317

Qy 287  EHNITADIEVIPDYINTAMERVKKDVRPFRFVIDVENT 325
Db 318  KHNITADIELIKMDQINTAMERLSKADVKYRFVIDVANS 356

```

RESULT 15
 US-10-424-599-205468
 ; Sequence 205468, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David X
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 205468
 ; LENGTH: 364
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_27566C.1.pgp
 US-10-424-599-205468

```

Query Match      47.9%; Score 842; DB 12; Length 364;
Best Local Similarity 49.6%; Pred. No. 1.9e-79;
Matches 171; Conservative 50; Mismatches 80; Indels 44; Gaps 6;

Qy 3 GATDVRFKVLGVGVCHSDHMAKNDGTSYPIVPGHSLVGVVTEVGCKVKKE----- 56
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 38 GMDNITLVLVYSGICHTLHMVNDPRISYIPWPQGEHIVGVTKVGRRTKSVGDIAG 97
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
QY 57 -SWQGRCLHGRLRPTCENCIHLENYCPNLQIOTYGSYYDGTMTYGGYSNNMVTDEHF 115
Db 98 VGGSVGSC-----GSCDCSNGFYVYCPWILTYSAHYDGTITQGGYSNNIWDQNF 150
QY 116 IVRIPDNLPDGAAPLLCAGITTYSPWYYGLDKPGMHLGVEMPRFRSRPPLNLPGLWG 175
Db 151 VWLIPKSLPLDGAAPLLCAGITTYSPWYYGLDKPGMHLGV-----VGLGG 196
QY 176 SRLQSLVPPLIKEGSY-----GTSPALMHS-----LRTDQDQMEAMSTMGGII 221
Db 197 --LGHVAVKFAKAFGMHVTWISTSPSKKEALEKLGADFEVSLDQOQLODARGTMDAIL 254
QY 222 DTVPVRLRPLISLLKTNKGWTVGINVQPLDLPVRLIIGRKWAGSAIGMKETOEM 281
Db 255 DTVSANHSIQPLIALLKTSGLIIVGGPPSPLEVLAMPLLGRKMIAGSAGGGRBIOEM 314
QY 282 IDFAAEHNITADIEVPIIDYLNATAMRVRVKKDVRFRFVIDVENTL 326
Db 315 MDPFAAKHNITADVEVPMVDYNTAFERLEKNDVKYRFVIDVANTI 359
```

Search completed: July 6, 2004, 13:48:42
Job time : 37.036 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:29:59 ; Search time 8.38046 Seconds
(without alignments)
2025.529 Million cell updates/sec

Title: US-09-857-518A-31

Perfect score: 1758

Sequence: 1 ERGATDVRFKVLYCGVCHSD.....ERVVKDVRFRVIDVENTIL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1092	62.1	359	1 MTD_FRAAN	Q92rf1 fragaria an
2	971	55.2	361	1 MTD_MESCR	P93257 mesembryant
3	963	54.8	337	1 MTD_PETCR	P42754 petroselinu
4	960	54.6	357	1 MTDI_ARATH	Q02971 arabidopsis
5	939	54.6	359	1 MTD_ARATH	Q82972 arabidopsis
6	946	53.8	365	1 MTD_APIGR	Q48707 apium grave
7	884	50.3	360	1 MTDH_ARATH	P42734 arabidopsis
8	870.5	49.5	359	1 MTD_MESPSA	O82515 medicago sa
9	860.5	48.9	363	1 MTD3_STYHU	Q43138 stylosanthe
10	833.5	47.4	354	1 MTDI_STYHU	O43137 stylosanthe
11	740	42.1	357	1 CAD2_PICAB	Q82035 picea abies
12	740	42.1	357	1 CAD7_PICAB	Q08350 picea abies
13	739	42.0	357	1 CADH_PINRA	P41637 pinus taeda
14	723	41.1	357	1 CADH_PINRA	Q40976 pinus radia
15	691.5	39.3	360	1 CAD1_ARACO	P42495 aralia cord
16	686	39.0	357	1 CAD9_TOBAC	P30360 nicotiana t
17	681	38.7	357	1 CAD4_TOBAC	P30359 nicotiana t
18	666	37.9	357	1 CADH_POPE	P31657 populus del
19	656	37.3	361	1 CADH_LOLEP	O22380 lolium pere
20	654	37.2	356	1 CAD2_EUCGU	P31655 eucalyptus
21	653	37.1	356	1 CADH_EUCGL	O64969 eucalyptus
22	646	36.7	357	1 CAD2_ARATH	Q49482 arabidopsis
23	641	36.5	358	1 CADH_MESDA	P31656 medicago sa
24	629	35.8	354	1 CAD1_EUCGU	Q42726 eucalyptus
25	628	35.7	367	1 CADH_MAIZE	O24562 zea mays (m
26	621.5	35.4	355	1 CADH_EUCBO	P50746 eucalyptus
27	621	35.3	365	1 CAD1_ARATH	P48523 arabidopsis
28	611	34.8	365	1 CADH_SACOF	O82056 saccharum o
29	599.5	34.1	346	1 ADH_MYCTU	P31975 mycobacteri
30	555.5	31.6	349	1 YAHK_ECOLI	P75691 escherichia
31	379	21.6	360	1 YW97_ECOLI	Q04894 saccharomyc
32	365.5	20.8	339	1 YJGB_ECOLI	P27250 escherichia
33	356.5	20.3	339	1 ADH3_BACST	P42328 bacillus st

RESULT 1

MTD_FRAAN	STANDARD;	PRT;	359 AA.
ID	MTD_FRAAN		
AC	Q92RF1;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Probable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase).		
GN	CAD.		
OS	Fragaria ananassa (Strawberry).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid1; Rosales; Rosaceae; Rosoideae; Fragaria.		
OX	NCBI_TaxID=3747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. Chandler;		
RA	Medina-Escobar N., Caballero J.L., Munoz-Blanco J.;		
RT	"Cloning, sequencing and temporal and spatial expression pattern of the cinnamyl alcohol dehydrogenase";		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level (By similarity).		
CC	-1- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.		
CC	-1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).		
CC	-1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.		
CC	-1- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol dehydrogenase.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U63534; RAD10327.1; --		
DR	InterPro; IPR002328; ADH_zinc.		
DR	InterPro; IPR002085; Adh_zn family.		
DR	Pfam; PF00107; ADH_zinc_N; --		
DR	PROSITE; PS00059; ADH_ZINC; 1.		
DR	KW Oxidoreductase; Zinc; Metal-binding; NAD.		
FT	METAL 48 48		
FT	METAL 70 70		
FT	METAL 101 101		
FT	METAL 104 104		
FT	METAL 107 107		
FT	METAL 115 115		
FT	METAL 164 164		
SQ	SEQUENCE 359 AA; 39137 MW; BAB1D6B3AF086DAB CRC64;		

ALIGNMENTS

34	350.5	19.9	337	1	ADH1_BACST	P12311	bacillus st
35	350.5	19.9	339	1	ADH2_BACST	P42327	bacillus st
36	320.5	18.2	361	1	YCZ5_YEAST	P25377	saccharomyc
37	316.5	18.0	340	1	ADH5_RHIME	O31186	rhizobium m
38	300.5	17.1	352	1	ADH3_EMENI	P07754	emeritella
39	294.5	16.8	348	1	ADH2_KLUJA	P49383	kluveromyc
40	285.5	16.2	348	1	ADH2_CANAL	O94038	candida alb
41	278	15.8	349	1	ADH1_ASFPL	P41747	aspergillus
42	277.5	15.8	350	1	ADH1_CANAL	P43067	candida alb
43	277.5	15.8	353	1	ADH1_NEUCR	Q9p6c8	neurospora
44	275.5	15.7	350	1	ADH1_KLUJA	P20369	kluveromyc
45	269.5	15.3	347	1	ADH2_KLUJA	Q9p4c2	kluveromyc

Query Match 62.1%; Score 1092; DB 1; Length 359;
 Best Local Similarity 65.7%; Pred. No. 2.7e-82;
 Matches 224; Conservative 27; Mismatches 56; Indels 32; Gaps 7;

QY 1 ETGATDVRFKVLYCGVCHSDIHMKNMGSTSTPIVPGHELGVVTEVCGKVKFK-SWR 59
 DB 32 ETGKDVTFKVLVYCGICSHLHMKNMGSTSTPIVPGHELGVVTEVCGKVKFKVGR 91

QY 60 QGRCLHGRLEPTCENCIHLENYCPNLIQYGSKYDGTMTYGGYNNMTDHFIVRI 113
 DB 92 VGVGCIWGSCR-SCENCTDHLNCPKOLLYGAKYDGTMTYGGYNNMTDHFIVRI 150

QY 120 PDNPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRFRSRPLNLPGLWGSRLQ 179
 DB 151 PDNPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRFRSRPLNLPGLWGSRLQ 194

QY 180 SLVPLIKEGG-----SVGTSP-----ALMH-----SLRTDODQMAAMSTMDGIIDTVP 225
 DB 195 HVAVKFAKAMGVKVTISTSPKXEEALHGLGADSFVSRDQDQMAAIGTMDGIIDTVP 254

QY 226 AVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRMVAGSAIGMKETQEMIDFA 285
 DB 255 AQHPLPLIGLNGSHGLVMVGAPEKPLELPPVPLIIGRMVAGSAIGMKETQEMIDFA 314

QY 286 AHNITADIEVPIIDYLTAMERVVKDVRFRFVIDVENTL 326
 DB 315 ARHNITADIEVPIIDYLTAMERVVKDVRFRFVIDENTL 355

RESULT 2

MTD_PETCR
 ID MTD_PETCR STANDARD; PRT; 361 AA.
 AC P42754;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase).
 GN ELI3.
 OS Mesembryanthemum crystallinum (Common ice plant).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Aizoaceae; Mesembryanthemum.
 OX NCBI_TaxID=3544;
 RN Michaelowski C.B., Bohnert H.J.;
 RP Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RL
 CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level (By similarity).
 CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.
 CC -!- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol dehydrogenase.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U79770; A0338503.1; -;
 PIR; T12571; T12571.
 InterPro; IPR002328; ADH_zinc.
 InterPro; IPR002085; Adh_zn family.
 Pfam; PF010107; ADH_zinc_N; 1.
 PROSITE; PS00059; ADH_ZINC; 1.

OX Oxidoreductase; Zinc; Metal-binding; NAD.
 FT METAL 51 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 73 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 104 ZINC 2 (BY SIMILARITY).
 FT METAL 107 ZINC 2 (BY SIMILARITY).
 FT METAL 110 ZINC 2 (BY SIMILARITY).
 FT METAL 118 ZINC 2 (BY SIMILARITY).
 FT METAL 167 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 361 AA; 39068 MW; 3BA2BFB946144D10 CRC64;

Query Match 55.2%; Score 971; DB 1; Length 361;
 Best Local Similarity 55.4%; Pred. No. 2.4e-72;
 Matches 190; Conservative 52; Mismatches 63; Indels 38; Gaps 6;

QY 2 TGATDVRFKVLYCGVCHSDIHMKNMGSTSTPIVPGHELGVVTEVCGKVKFK-SWRQ 60
 DB 36 TGEDQVTFKVLVYCGICSHLHMKNMGSTSTPIVPGHELGVVTEVCGKVKFKVGR 95

QY 61 GRCMLHGRLEPTCENCIHLENYCPNLIQYGSKYDGTMTYGGYNNMTDHFIVRI 120
 DB 96 GVGCMVGSRC-SCSCENHLENYCPNLIQYGSKYDGTMTYGGYNNMTDHFIVRI 154

QY 121 DNPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRFRSRPLNLPGLWGSRLQ 180
 DB 155 DNMALDAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRFRSRPLNLPGLWGSRLQ 195

QY 181 LVPLLIKEGGSYG-----TSPALMHS-----LLRTDQDQMAAMSTMDGIIDT 223
 DB 196 LGHMAVFKGAKGVKVTISTSPKXEEALHGLGADSFVSRDQDQMAAIGTMDGIIDT 255

QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRMVAGSAIGMKETQEMID 283
 DB 256 VSAHPLPLIGLNGSHGLVMVGAPEKPLELPPVPLIIGRMVAGSAIGMKETQEMID 315

QY 284 PAAHNITADIEVPIIDYLTAMERVVKDVRFRFVIDVENTL 326
 DB 316 PAAKHIDKSDIEVVPMDYVNTAMERLLKGVRYRFRVIDVANTL 358

RESULT 3

MTD_PETCR
 ID MTD_PETCR STANDARD; PRT; 337 AA.
 AC P42754;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase) (Fragment).
 GN ELI3.
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;
 OC Apium clade; Petroselinum.
 OX NCBI_TaxID=4043;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93099840; PubMed=1464303;
 RA Kiedrowski S., Kawalleck P., Hahlbrock K., Somssich I.E., Dangel J.L.;
 RT "Rapid activation of a novel plant defense gene is strictly dependent on the Arabidopsis RPM1 disease resistance locus";
 RL EMBO J. 11:4677-4684(1992).
 CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level (By similarity).
 CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X67817; CAA48028.1; --
 PIR; S28045; S28045; ADH zinc.
 InterPro; IPR002328; ADH zinc.
 InterPro; IPR002085; Adh zn family.
 Pfam; PF00107; ADH zinc N; 1.
 DR PROSITE; P500059; ADH ZINC; 1.
 KW Oxidoreductase; Zinc; Metal-binding; NAD.
 FT NON_TER 1
 FT METAL 27 27 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 49 49 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 80 80 ZINC 2 (BY SIMILARITY).
 FT METAL 83 83 ZINC 2 (BY SIMILARITY).
 FT METAL 86 86 ZINC 2 (BY SIMILARITY).
 FT METAL 94 94 ZINC 2 (BY SIMILARITY).
 FT METAL 143 143 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 337 AA; 36209 MW; 8C71A08D8CAD5656 CRC64;
 Query Match 54.8%; Score 963; DB 1; Length 337;
 Best Local Similarity 55.2%; Pred. No. 9.9e-72;
 Matches 191; Conservative 46; Mismatches 65; Indels 44; Gaps 5;
 QY 2 TGATVRFKLVGVGVCHSHMAKNDWGTSTPIVPGHVLGVTVIVGCKVKKF----- 56
 DB 12 TGDNVRFKLVGVGVCHSHMAKNDWGTSTPIVPGHVLGVTVIVGCKVKKF----- 71
 QY 57 --SWQGRCLHGRLEPTCENCIHLENYCPNLIOTYGSKYVDGTYGYSNNWYTDH 114
 DB 72 GVGCLVGSCL-----SCNCDSDSNNCAKQVQYAFNVDGSIYGGYADSMVADQH 124
 QY 115 FIVRIPLDPLDCAAPLLCAGITTYSPWRYGKDEKPMHGVWVPRFRSRPLNLPGLW 174
 DB 125 FVLRWPNLPLDGAAPLLCAGITTYSPWRYGKDEKPMHGVWVPRFRSRPLNLPGLW 170
 QY 175 GRLSLVPLIKEGSYGTSALMHS-----LLRTDQDQMEAMSTMDGI 220
 DB 171 G--LGHVAVKAKAFGAHVTVISTSSKQKQKLEKLGADEFVLVSDSDQQAATGLHGI 228
 QY 221 IDTVPAVPLEPLISLTKNGKVTVTVGIAVQPLDLPVPLIIGKRWAGSAIGMKETQE 280
 DB 229 IDTVSAHLPVPLGLKNGKLVWVGAPEKPELPELFPFLMGRKRVAGSNIGGLASTQE 288
 QY 281 MIDFAAEHNITADISVIPIDYLTAMERVVKDVRFRFVIDVENTL 326
 DB 289 MLDFAAQENITADVIVPDYVNTAMERLVKSDVRFRFVIDVANTI 334

RESULT 4

MTD1 ARATH
 ID MTD1 ARATH STANDARD; PRT; 357 AA.
 AC Q02971; Q9SZJ9;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Probable mannitol dehydrogenase 1 (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase 1).
 GN BL13-1 OR AT4G37980 OR P20D10.100.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=93099840; PubMed=1464303;
 RA Kiedrowski S., Kwallack P., Hahlbrock K., Somssich I.E., Dangl J.L.;
 RT "Rapid activation of a novel plant defense gene is strictly dependent on the Arabidopsis RPM1 disease resistance locus.";
 EMBO J. 11:4677-4684 (1992).
 [2]
 RT SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entwistle K.-D., Terryn N., Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M., Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M., Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T., Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hobeisel J., Zimmermann W., Medler H., Ridley P., Langham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grynprenz B., Chuang Y.-J., Vandenbussche F., Braeken M., Weitzens I., Voet M., Bastiaens I., Aert R., Defoor E., Weitzenecker T., Bothe G., Rampsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirks W., Mooliman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Berneiser S., Hempel S., Feldpausch R., Lamberth S., Van den Daele H., De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lemard N., McLeay K., Mayes R., Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S., Borkova D., Blacecker H., Scharfe M., Grizm M., Loehner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E., Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chedfor G., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A., Perez-Perez A., Purnelle S., Bent E., Johnson S., Tacón D., Jese T., Heishman L., Schwarz S., Scholler P., Heber S., Strasser P., Bielke C., Feinman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegele L., Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J., Stoneking T., Kallio J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W., Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Bergthoff A., Jones K., Drone K., Cotton M., Joshua C., Antonio B., Zidani B., Strong C., Sun H., Lamar B., Yordan C., Ma P., Zhong J., Preston R., Vil D., Shekter M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M.A., Martienssen R., McCombie W.R.;
 "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";
 Nature 402:769-777(1999).
 RL Nature 402:769-777(1999).
 CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level (By similarity).
 CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR InterPro: IPR002328; ADH zinc.
DR Pfam: PF00107; ADH zinc N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 68 68 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 102 102 ZINC 2 (BY SIMILARITY).
FT METAL 105 105 ZINC 2 (BY SIMILARITY).
FT METAL 113 113 ZINC 2 (BY SIMILARITY).
FT METAL 162 162 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 6 6 E -> Q (IN REF. 1).
FT CONFLICT 17 17 E -> N (IN REF. 1).
FT CONFLICT 20 20 I -> V (IN REF. 1).
SQ SEQUENCE 357 AA; 38245 MW; 2C27B3C2BF030166 CRC64;

Query Match 54.6%; Score 960; DB 1; Length 357;
Best Local Similarity 55.4%; Pred. No. 1.9e-71;
Matches 190; Conservative 47; Mismatches 68; Indels 38; Gaps 5;

QY 2 TGATDVPRKVLGVGCHSDIIMAKNDWGTSTPIVPGHVLGVWTEVGVKKKFSW-RQ 60
DB 31 TGEKDVPRKVLGVGCHSDIIMAKNDWGTSTPIVPGHVLGVWTEVGVKKKFSW-RQ 90
QY 61 GRCWLHGLRPTCEHCIIHLENYCNLIQTYGSKYDGTMTYGGYSNNWVTDEHFIVRIP 120
DB 91 GVGVMAGSCR-SCDCNDGDENYCPKMLITSGAKNFDFTWTHGGYSDEWVCAEDFIIRIP 149
QY 121 DNLPLDGAAPLLCAGITYSPWYGLDKPGMHLGVWEPFRPRPPLNLPOLWSRLQS 180
DB 150 DNLPLDGAAPLLCAGITYSPWYGLDKPGMHLGVWEPFRPRPPLNLPOLWSRLQS 190
QY 181 LVPLLEKGGVGTSPALMHS-----LLRTDDQDQAAAMSTMDGILDT 223
DB 191 LCHVAVFPAKANGTQVTVISERKRDVAETLGDADFVSRDPKMDAMGTMDGILDT 250
QY 224 VPAVRPLEPLSLAKTNGKVVTVGVIAVQPLDPVPLLIIGRMVAGSAGGKMKETQEMID 283
DB 251 VSATHPLPLGLLKNKGLKLVGAPAPLEPLVPFLIFGRKMWVGVSGVIGKIKETQEMVD 310
QY 284 FAEENHTADIEVPIIDVINTAMERVKVQVFRFVIDVENTL 326
DB 311 LAGKENTADIELISADYVNTAMERLAKADVKEFVIDVANTM 353

RESULT 5
MTD2 ARATH STANDARD; PRT; 359 AA.
AC Q02972;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable mannitol dehydrogenase 2 (EC 1.1.1.255) (NAD-dependent
DE mannitol dehydrogenase 2).
GN ELI3-2 OR AT4G37990 OR F20D10.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93099840; PubMed=1464303;
RA Kiedrowski S., Kawalleck P., Hahlbrock K., Somssich I.E., Dangl J.L.;
RT "Rapid activation of a novel plant defense gene is strictly dependent
RL on the Arabidopsis RPM1 disease resistance locus.";
RL EMOB J. 11:4677-4684(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083486; PubMed=10617198;

```

```

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Voickaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoerge W., Brandt P., Grivelli L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier S., Mache R., Mueller K.,
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Woltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzreger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer B., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., Mclay K., Mayes R.,
RA Borkova D., Bloeker H., Scharfe M., Benes V., Rechmann S.,
RA Dose S., de Haan M., Maarse A.C., Schaefer K., Loehner T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedors T., Weber N., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Cooke R., Vandenbol M., Barques M., Terol J., Torres A.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney E., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step
CC by which translocated mannitol is committed to central metabolism
CC and, by regulating mannitol pool size, is important in regulating
CC salt tolerance at the cellular level (By similarity).
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; X67815; CAB48026.1; -
EMBL; AL035538; CAB37539.1; -
EMBL; AL161592; CAB80464.1; -
PIR; S28043; S28043.
InterPro; IPR002328; ADH_zinc.
Pfam; PF00107; ADH_zinc_N; 1.
PROSITE; PS00059; ADH_ZINC; 1.
Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
METAL 46 46 ZINC 1 (CATALYTIC) (BY SIMILARITY).

```

FT METAL 68 68 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 99 99 ZINC 2 (BY SIMILARITY).
FT METAL 102 102 ZINC 2 (BY SIMILARITY).
FT METAL 105 105 ZINC 2 (BY SIMILARITY).
FT METAL 113 113 ZINC 2 (BY SIMILARITY).
FT METAL 162 162 ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 359 AA; 38942 MW; B691F988AD4842A5 CRC64;

Query Match 54.6%; Score 959; DB 1; Length 359;
Best Local Similarity 55.1%; Pred. No. 2.3e-71;
Matches 193; Conservative 39; Mismatches 68; Indels 50; Gaps 5;

QY 1 ETGATDVRKVLGVCHSDIHMKNKDWGTSYPIVPGHELVGVTVGCKVKKSWRQ 60
DB 30 ETGEKDVRFKVLGCGCHSDIHMKNKDWGTSYPIVPGHELVGVTVGCKVKKSWRQ 89
QY 61 -----GRCWLHGRLEPFCNCIHLHLENCPNLIQTYGSKYDGTWYGYNNMVYDE 113
DB 90 VGVGCLVSSC-----GSCDSCTEGMEYCPKSIQTYGFPYDNTITYGYSDHVVCEE 142
QY 114 HPTVIRPDNPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRSRPLNPLGL 173
DB 143 GFVIRPDNPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRSRPLNPLGL 188
QY 174 WGSRLQSLVPLIKEGSSYGTSPALMHS-----LLRTDQDQMEAAAMST 216
DB 189 GG-----LGHVGVKFAKAMGKVTVIISTSEKKRDEALNRLGADAFVSRDPKQKIDMGT 243
QY 217 MDGIIDTVPVAVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVPLIIGRKNVAGSAIGMK 276
DB 244 MDGIIDTVPVAVRPLEPLISLKTNGKVVTVGVIAVQPLDLPVPLIIGRKNVAGSAIGMK 303
QY 277 ETQEMIDFAEHNITADIEVPIIDYNTAMERVVKKVVRFRFVIDVENTL 326
DB 304 ETQEMIDMAGKNITADIEVPIIDYNTAMERVVKKVVRFRFVIDVENTL 353

RESULT 6

MTD APIGR STANDARD; PRT; 365 AA.
AC Q38707; O82461;
DT 16-OCT-2001 (Rel. 40, Created).
DT 16-OCT-2001 (Rel. 40, Last sequence update).
DT 10-OCT-2003 (Rel. 42, Last annotation update).
DE Mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase).
GN MTD.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apicoideae; apioid superclade;
OC Apium clade; Apium.
OX NCBI_TaxID=4045;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95365325; PubMed=7638158;
RA Williamson J.D., Stoop J.M.H., Massel M.O., Conkling M.A., Pharr D.M.;
RT "Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein EL13.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:7148-7152 (1995).
RN [2]
RN SEQUENCE FROM N.A.
RA Williamson J.D., Guo W.-W., Pharr D.M.;
RT "Cloning and characterization of a genomic clone encoding mannitol dehydrogenase from celery (Apium graveolens).";
RL (In) Plant Gene Register PGR98-137.
RN [3]
RN CHARACTERIZATION
RA Stoop J.M.H., Chilton W.S., Pharr D.M.;
RT "Substrate specificity of the NAD-dependent mannitol dehydrogenase from celery.";
RL Phytochemistry 43:1145-1150 (1996).
CC -i- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step

by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level.
-i- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
-i- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
-i- SUBCELLULAR LOCATION: Cytoplasmic.
-i- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U24561; AAC15467.1; --
CC EMBL; AF067082; AAC61854.1; --
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn_family.
CC InterPro; IPR000205; NAD_BS.
CC InterPro; IPR000051; SAM_Bind.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD.
KW METAL 50 50 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 72 72 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 109 109 ZINC 2 (BY SIMILARITY).
FT METAL 117 117 ZINC 2 (BY SIMILARITY).
FT METAL 166 166 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 47 47 C -> S (IN REF. 2).
FT CONFLICT 120 120 T -> I (IN REF. 2).
SQ SEQUENCE 365 AA; 39690 MW; A46E394F5DCB077F CRC64;

Query Match

Best Local Similarity 53.8%; Score 946; DB 1; Length 365;
Matches 192; Conservative 42; Mismatches 74; Indels 32; Gaps 5;

QY 2 TGATDVRKVLGVCHSDIHMKNKDWGTSYPIVPGHELVGVTVGCKVKKE-SWRQ 60
DB 35 TGEKDVRLKVLGCGCHSDIHMKNKDWGTSYPIVPGHELVGVTVGCKVKKEGVNV 94
QY 61 GRCWLHGRLEPFCNCIHLHLENCPNLIQTYGSKYDGTWYGYNNMVTDHPIVRIP 120
DB 95 GIGCLVSGCR-SCSCCNDRESHCENTIDTYGSIYFDGTWCHGGYSTMTWADHFIILRW 153
QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRSRPLNPLGLSGRLQS 180
DB 154 KNLPLDGAAPLLCAGITTYSPWRYGLDKPGMHLGVNPPRRSRPLNPLGLSGRLQS 197
QY 181 LVPLIKEGSSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDTVPV 226
DB 198 VAVKMAKAFAGQWTVIDISESKRKEALEKLGADFLNSDQEQMGARSLDGIIDTVPV 257
QY 227 VRPLEPLISLKTNGKVVTVGVIAVQPLDLPVPLIIGRKNVAGSAIGMKETQEMIDFAA 286
DB 258 NHPLAPLPDLKPKNGKLVWGAPEKPELFPVSLKGRKLGTTGGTGGKKEKTOEMLDFAA 317
QY 287 EHNITADIEVPIIDYNTAMERVVKKVVRFRFVIDVENTL 326
DB 318 KHNITADIEVPIIDYNTAMERVVKKVVRFRFVIDIANTM 357

RESULT 7

MTD ARATH STANDARD; PRT; 360 AA.
AC P42734;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

Protable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase).

CAD1 OR AT4G39330 OR T22F8.230.

Arabidopsis thaliana (Mouse-ear cress).

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

MEDLINE=95357425; PubMed=7610954;

Scmex D.A., Mourse J.P., Manners J.M., Abrahams S.L., Watson J.M.;

A gene encoding a cinnamyl alcohol dehydrogenase homolog in Arabidopsis thaliana.

Plant Physiol. 108:1309-1310(1995).

[2]

SEQUENCE FROM N.A.

STRAIN=cv. Columbia;

MEDLINE=20083488; PubMed=10617198;

Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N., Harris B., Ansgore W., Brandt P., Grivell L.A., Rieger M., Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller T., Kreis M., Delsen Y.M., Puigdomenech P., Watson M., Schmidheini T., Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P., Langham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Meltsens I., Voet M., Bastiaens I., Aert R., Defoor E., Weizenecker T., Bothe G., Kampsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R., Pettett A., Rajadream H.A., Lyne M., Benes V., Reckmann S., Borikova D., Bloecker H., Schaffe M., Grimm M., Loehnert T.-H., Dose S., de Haan M., Maare A.C., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A., Neumann S., Argiriou A., Vitale D., Lignori R., Piravandi E., Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt M., Lecharny A., Aubourg S., Cherdorf F., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T., Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Biele C., Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dehnan N., Gnoj L., Schutz K., Huang E., Spiegel L., Sekonk M., Murray J., Sheet P., Cordes M., Abu-Threideh J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Winx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kraner J., Fulton L., Mardis E., Danté M., Pepin K., Hillier L.W., Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C., Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C., Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M.A., Martensen R., McCombie W.R.;

Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

Nature 402:769-777(1999).

-!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step by which translocated mannitol is committed to central metabolism and, by regulating mannitol pool size, is important in regulating salt tolerance at the cellular level (By similarity).

-!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.

-!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).

-!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase family.

-!- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol dehydrogenase.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L37883; AAA99511.1; -

EMBL; L37884; AAA74746.1; -

EMBL; AL050351; CAB43648.1; -

EMBL; AL161595; CAB80596.1; -

InterPro; IPR002328; ADH_zinc.

InterPro; IPR002085; Adh_zn family.

Fam; PF00107; ADH_zinc_N; 1.

PROSITE; P580059; ADH_ZINC; 1.

KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.

FT METAL 50 50 ZINC 1 (CATALYTIC) (BY SIMILARITY).

FT METAL 72 72 ZINC 1 (CATALYTIC) (BY SIMILARITY).

FT METAL 103 103 ZINC 2 (BY SIMILARITY).

FT METAL 106 106 ZINC 2 (BY SIMILARITY).

FT METAL 109 109 ZINC 2 (BY SIMILARITY).

FT METAL 117 117 ZINC 2 (BY SIMILARITY).

FT METAL 166 166 ZINC 1 (CATALYTIC) (BY SIMILARITY).

FT VARIANT 202 202 S -> I.

SQ SEQUENCE 360 AA; 38907 MW; 82C270E595C713A7 CRC64;

Query Match 50.3%; Score 884; DB 1; Length 360;

Best Local Similarity 49.1%; Pred. No. 3.3e-65;

Matches 172; Conservative 51; Mismatches 77; Indels 50; Gaps 5;

QY 1 ETGATDVRFKLYCGVCHSDIHMAKNDWGSTYDIPVGHVGVTVGCKVKKFKSWRQ 60

DB 34 DNGENDVTVKILFCGVCHTDLHTINDWGSYYYPWFGHEIVGIATKVGKNTKFKEGDR 93

QY 61 -----GRCVHLGRRLPTCNCIHLNENYCPNLIQYGYKGYDGTWYGYGNMWTDE 113

DB 94 VCVGVVISGC-----QSCSCQDLENYCPQSFYNAIGSDGTNKGYSNIVVDQ 146

QY 114 HFIVIPONLPDGAAPLLCAGITTYSPWRYGDLKPGMELGVENPFRSRPLNLEGL 173

DB 147 RVLRFPPNLESDSGAPLLCAGITVYSPMKYGYMGTEAGKELGV-----AGL 192

QY 174 WGRSLQSLVPLIKEGSYGTSPLMH-----LRTDQDMERAKST 216

DB 193 GG-----LGRVAVSGKAGKLVTVISSSTKASEA.NHLGADSFLLTDPQKKAIGT 247

QY 217 MDGIIDTPVAPRPLEPLISILKTKGVTVGVIAVQPLDLPFPLIGRKNVAGSAGGMK 276

DB 248 MDYIIDTTSVAVHALYPLGLLKVKNGKLIAGLPEKPLEPMLPVLGRKNVGSVDGGMK 307

QY 277 ETQEMIDFAEHNITADIEVIPDYNATAXERVVKQDVRFRFVIDVENTL 326

DB 308 ETQEMLDFCAHNITADIELIKMDEINTAXERLAKSDVRYRVIDVANSL 357

RESULT 8

MTD MEDSA

ID MTD MEDSA STANDARD; PRT; 359 AA.

AC 082515;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable mannitol dehydrogenase (EC 1.1.1.255) (NAD-dependent mannitol dehydrogenase).

GN CAD1.

OS Medicago sativa (Alfalfa).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; OC

```

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=cv. Siriver; TISSUE=stem;
RX MEDLINE=20044095; PubMed=10579494;
RA Brill E.M., Abrahams S., Hayes C.M., Jenkins C.L., Watson J.M.;
RT "Molecular characterization and expression of a wound-inducible cDNA
  encoding a novel cinnamyl-alcohol dehydrogenase enzyme in lucerne
  (Medicago sativa L.)";
RL Plant Mol. Biol. 41:279-293(1999).
CC -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step
  by which translocated mannitol is committed to central metabolism
  and, by regulating mannitol pool size, is important in regulating
  salt tolerance at the cellular level (By similarity).
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
  family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; AF083333; AAC35846.1; -.
CC InterPro; IPR002328; ADH_zinc.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD.
CC METAL 50 50 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 72 72 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 103 103 ZINC 2 (BY SIMILARITY).
CC METAL 106 106 ZINC 2 (BY SIMILARITY).
CC METAL 109 109 ZINC 2 (BY SIMILARITY).
CC METAL 117 117 ZINC 2 (BY SIMILARITY).
CC METAL 165 165 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 359 AA; 39027 MW; D84E383E0F195A1 CRC64;
CC
CC Query Match 49.5%; Score 870.5; DB 1; Length 359;
CC Best Local Similarity 52.0%; Pred. No. 4.2e-64;
CC Matches 178; Conservative 42; Mismatches 85; Indels 37; Gaps 5;
CC
CC QY 1 ETGATDVRPKVLYCGVCHSDIHMANKDWTSTYTPVPGHELGVVTVGVCKVKFKSWRQ 60
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC 34 ENGDDVSVKILYCGVCHSDLTLDKDWGFTTYPVPGHEIVGVTVGVINVKKFRVGDN 93
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 61 GRCLWHLRLPTCNCIHLNENYCNLIQTYGSKYDGTMTYGGYSNNMVTDEHFIVIP 120
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 94 VGVGVIVRESQTCNCNQLDLEQCEKFPVFTNSP-YKGRTRYGGYDFVGVHQRVYVQFP 152
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGHGLGVWPRFRSRPPLNPLGLWSRLQS 180
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 153 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGHGLGVWPRFRSRPPLNPLGLWSRLQS 193
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 181 LVPPPLIKEGSGY-----TSP-----ALMHSLLRTDQDQEAAMSTMDGIIDT 223
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 194 LGHVAIKFGAKFLKVTIVSTSPNKETEADIKLGADSLVSKDPEKMAAGTMDYIIDT 253
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 224 VPAVRPLEPLSLKLTNGKVTGVGLVAPLDLPPLPPLIIGRWAGSAIGCKKETEIMD 283
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 254 ISAAHSLAPLGLKGLKLVTVGLPSKPLSLVFPPLVAGRLIGGSGNIGCKKETEIMD 313
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 284 FAAGHNTADLEVPIDVINTAMERVVKDKVFRFVIDVENT 325
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 314 FCGKHNTADLELVHINTEINAMERLKHADVKYRFVIDVANS 355
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 9

```

MTD3 STYHU
ID MTD3 STYHU STANDARD; PRT; 363 AA.
AC Q43138;1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Probable mannitol dehydrogenase 3 (EC 1.1.1.255) (NAD-dependent
  mannitol dehydrogenase 3).
GN CAD3.
OS Stylosanthes humilis (Townselle stylo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Stylosanthes.
OX NCBI_TaxID=35628;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Paterson; TISSUE=stem;
RC Nourse J.P., Manners J.M., Curtis M.D., Abrahams S.L., Watson J.M.;
  Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: Oxidizes mannitol to mannose. Provides the initial step
  by which translocated mannitol is committed to central metabolism
  and, by regulating mannitol pool size, is important in regulating
  salt tolerance at the cellular level (By similarity).
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-mannose + NADH.
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
  family.
CC -!- CAUTION: Was originally (Ref.1) thought to be a cinnamyl-alcohol
  dehydrogenase.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; L36456; AAA74883.1; -.
CC InterPro; IPR002328; ADH_zinc.
CC InterPro; IPR002085; Adh_zn family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
CC METAL 51 51 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 73 73 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC METAL 104 104 ZINC 2 (BY SIMILARITY).
CC METAL 107 107 ZINC 2 (BY SIMILARITY).
CC METAL 110 110 ZINC 2 (BY SIMILARITY).
CC METAL 118 118 ZINC 2 (BY SIMILARITY).
CC METAL 168 168 ZINC 1 (CATALYTIC) (BY SIMILARITY).
CC SEQUENCE 363 AA; 39551 MW; E8EB217274513D90 CRC64;
CC
CC Query Match 48.9%; Score 860.5; DB 1; Length 363;
CC Best Local Similarity 50.3%; Pred. No. 2.8e-63;
CC Matches 172; Conservative 52; Mismatches 85; Indels 33; Gaps 6;
CC
CC QY 1 ETGATDVRPKVLYCGVCHSDIHMANKDWTSTYTPVPGHELGVVTVGVCKVKFKSWRQ 60
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 35 DIGEDVALEVLVYCGICHTDLFMAKDFGNSIYPVPGHEIVGVTVGVKVKYKVGDK 94
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 61 GRCLWHLRLPTCNCIHLNENYCNLIQTYGSKYDGTMTYGGYSNNMVTDEHFIVIP 120
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 95 VGVGVIVRESQTCNCNQLDLEQCEKFPVFTNSP-YKGRTRYGGYDFVGVHQRVYVQFP 154
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPGHGLGVWPRFRSRPPLNPLGLWSRLQ 179
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 155 BGLPLDGGSSSLCWGY-SHSPKTYGLDKPGLHGV-----VGLGG--LG 197
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC QY 180 SLVPLPIKEGG-----SYGTSPALMHS-----LLRTDQDQEAAMSTMDGIIDT 224
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
CC DB 198 HVAKFAKTHGLKITVISTSPPTKKEAIKNLGADSLVSRDPQMEAPKETLDGIIDTV 257
CC :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Query Match 47.4%; Score 833.5; DB 1; Length 354;
Best Local Similarity 49.0%; Pred. No. 4.5e-61;
Matches 170; Conservative 47; Mismatches 79; Indels 51; Gaps 7.

FT METAL 114 114 ZINC 2 (BY SIMILARITY).
 SQ SEQUENCE 357 AA; 38834 MW; 56P81339B9E3097FC CRC64;
 Query Match 42.1%; Score 740; DB 1; Length 357;
 Best Local Similarity 42.7%; Pred. No. 2.1e-53;
 Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKLYCGVCHSDIHMANKDGTSTYPIVPGHELVGVTEVGCCKVKKESMRQ-- 60
 DB 33 GPDVTVIRVIYCGICHSDDLVQWGENEMSNYPWPVPGHEVGVTEIGSEVKKFKVGHVG 92
 QY 61 -----GRCMLHGLRLPTCNCIHLNENYCPNLITQYSGKYDGTMTYGGYNNMTDHEH 115
 DB 93 VGCIVGSC-----RSCSNCNGSMGQYCSKRIWTYNDVNHDTPTQGGFASWVVDQMF 145
 QY 116 IVRIPNLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPPLNPLGLWG 175
 DB 146 VVRIPENPLEQAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPPLNPLGLWG 194
 QY 176 SRLQSLVPLLIKGGSGVGTSPALMHS-----LLRTDQDQMEAAAMSTMD 218
 DB 195 MG-----VKIAFAFLHVTVISSDKKKERALEVLGADAYLVSKDAEKMQEASLD 246
 QY 219 GIIDTVPVPRPLEPLSLKTKNGKVTYVGIAGVQPLDLPVPLIIGRMVAGSAIGMKET 278
 DB 247 YIMDTIPVAHPLEPLALTKNGKLVMLGVVPEPLHFTVPLLLILGRSISAGSFIGSMEET 306
 QY 279 QEMIDFAAHNITADIEVIDYNTAMERVKVKKDVRFRVIDV 322
 DB 307 QETLDFCAKKVSSMIEVGLDYINTAMERLVKNDVRYRFVVDV 350

RESULT 12
 CAD7 PICAB STANDARD; PRT; 357 AA.
 ID Q08350;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cinnamyl-alcohol dehydrogenase 7/8 (EC 1.1.1.195) (CAD 7/8).
 GN CAD7 AND CAD8.
 OS Picea abies (Norway spruce) (Picea excelsa).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
 OX NCBI_TaxID=3329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94033286; PubMed=8219046;
 RA Galliano H., Cabane M., Beckerskorn C., Lottspeich F.,
 RA Sandermann H. Jr., Ernst D.;
 RT "Molecular cloning, sequence analysis and elicitor/ozone-induced
 accumulation of cinnamyl alcohol dehydrogenase from Norway spruce
 (Picea abies L.)."
 RT Plant Mol. Biol. 23:145-156(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Schubert R., Sperisen C., Mueller-Starck G., La Scala S., Ernst D.,
 RA Sandermann H. Jr., Haeger K.-P.;
 RT "The cinnamyl alcohol dehydrogenase gene family in Picea abies (L.)
 and phylogenetic relationships."
 RL Trees 12:453-463(1998).
 CC -!- FUNCTION: This protein catalyzes the final step in a branch of
 CC phenylpropanoid synthesis specific for production of lignin
 CC monomers. It acts on coniferyl-, sinapyl-, 4-coumaryl- and
 CC cinnamyl-alcohol.
 CC CATALYTIC ACTIVITY: Cinnamyl alcohol + NADP(+) = cinnamaldehyde +
 CC NADPH.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- PATHWAY: Lignin biosynthesis.
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 CC family.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X72675; CAA51226.1; --
 DR EMBL; AJ001925; CAA05096.1; --
 DR EMBL; AJ001926; CAA05097.1; --
 DR PIR; S39509; S39509.
 DR InterPro; IPR002328; ADH_zinc.
 DR Pfam; PF00107; ADH_zinc_N; I.
 DR PROSITE; PS00059; ADH_ZINC; I.
 KW Oxidoreductase; Zinc; Metal-binding; NADP; Lignin biosynthesis;
 KW Multigene family.
 FT METAL 47 47 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 69 69 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 100 100 ZINC 2 (BY SIMILARITY).
 FT METAL 103 103 ZINC 2 (BY SIMILARITY).
 FT METAL 106 106 ZINC 2 (BY SIMILARITY).
 FT METAL 114 114 ZINC 2 (BY SIMILARITY).
 FT METAL 163 163 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 357 AA; 38777 MW; 8CD26BC41B87CA92 CRC64;
 Query Match 42.1%; Score 740; DB 1; Length 357;
 Best Local Similarity 42.7%; Pred. No. 2.1e-53;
 Matches 147; Conservative 61; Mismatches 86; Indels 50; Gaps 5;

QY 3 GATDVRFKLYCGVCHSDIHMANKDGTSTYPIVPGHELVGVTEVGCCKVKKESMRQ-- 60
 DB 33 GPDVTVIRVIYCGICHSDDLVQWGENEMSNYPWPVPGHEVGVTEIGSEVKKFKVGHVG 92
 QY 61 -----GRCMLHGLRLPTCNCIHLNENYCPNLITQYSGKYDGTMTYGGYNNMTDHEH 115
 DB 93 VGCIVGSC-----RSCSNCNGSMGQYCSKRIWTYNDVNHDTPTQGGFASWVVDQMF 145
 QY 116 IVRIPNLDGAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPPLNPLGLWG 175
 DB 146 VVRIPENPLEQAAPLLCAGITTYSPWRYGLDKPMHGLGVWEPFRFRPPPLNPLGLWG 194
 QY 176 SRLQSLVPLLIKGGSGVGTSPALMHS-----LLRTDQDQMEAAAMSTMD 218
 DB 195 MG-----VKIAFAFLHVTVISSDKKKERALEVLGADAYLVSKDAEKMQEASLD 246
 QY 219 GIIDTVPVPRPLEPLSLKTKNGKVTYVGIAGVQPLDLPVPLIIGRMVAGSAIGMKET 278
 DB 247 YIMDTIPVAHPLEPLALTKNGKLVMLGVVPEPLHFTVPLLLILGRSISAGSFIGSMEET 306
 QY 279 QEMIDFAAHNITADIEVIDYNTAMERVKVKKDVRFRVIDV 322
 DB 307 QETLDFCAKKVSSMIEVGLDYINTAMERLVKNDVRYRFVVDV 350

RESULT 13
 CADH_PINTA STANDARD; PRT; 357 AA.
 ID CADH_PINTA
 AC P41637;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) (CAD).
 OS Pinus taeda (Loblolly pine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
 OX NCBI_TaxID=3352;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Xylem;
 RX MEDLINE=95327049; PubMed=7603432;


```

QY 61 -----GRCWLHGRURPTCENCIIHLENYCPNLIOTYGSKYVDGTMVYGGYSNNMVVDHFP 115
Db 93 VGVITVGC-----RSCGNCNQSQMEQYCSKRIWTYNDVNDGTPQTGGFASSMVVDQMF 145
QY 116 IVRIPNLPLDGAAPLLCAGITTYSPWRYYGLDKFGMHLGVWPRFRSRPLNLPGLWG 175
Db 146 VVRIPELPLDGAAPLLCAGITTYSPWRYYGLDKFGMHLGVWPRFRSRPLNLPGLWG 194
QY 176 SRLQSLVPLPLKEGSGYTSALMHS-----LRTDQDOMEAAMSTMD 218
Db 195 MG-----VKIAKAFGLHVTIVSSDKKKEAMEVLGADAYLVSKDTOMERAEASLD 246
QY 219 GIIDTPAVRPLEPLISLKTNGKVVTVGIAVQPLDLVPFLLITGRKMVAGSAIGMKET 278
Db 247 YIMDTIPVAHLEPLLYALLKNGKLMVGVPELHFVTPLLILGRSIAGSFSGSMEET 306
QY 279 QEMIDFAAENHITADIEVIPIDYLTAMERVVKXDVFRFVIDV 322
Db 307 QETLDFCAKKVSMIEVVGLDYINTAMERLEKNDVRYRFVVDV 350

RESULT 15
CADI_ARACO
ID CADI_ARACO STANDARD; PRT; 360 AA.
AC P42495;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cinnamyl-alcohol dehydrogenase 1 (EC 1.1.1.195) (CAD).
GN CADI.
OS Aralia cordata (Udo) (Cordate spikenard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Apiales; Araliaceae; Aralia.
OX NCBI_TaxID=29746;
RN [1]
RP SEQUENCE FROM N.A.
RA Hibino T., Shibata D., Chen J.-O., Higuchi T.;
RT "Cinnamyl alcohol dehydrogenase from Aralia cordata: cloning of the
RT cDNA and expression of the gene in lignified tissues.";
RL Plant Cell Physiol. 34:659-665(1993).
CC -1- FUNCTION: This protein catalyzes the final step in a branch of
CC phenylpropanoid synthesis specific for production of lignin
CC monomers. It acts on coniferyl-, sinapyl-, 4-coumaryl- and
CC cinnamyl-alcohol.
CC -1- CATALYTIC ACTIVITY: Cinnamyl alcohol + NADP(+) = cinnamaldehyde +
CC NADPH.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Lignin biosynthesis.
CC -1- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----EMBL; D13991; BAA0309.1; --
CC InterPro; IPR002328; ADH zinc.
CC InterPro; IPR002085; Adh_zn family.
CC Pfam; PF00107; ADH_zinc_N; 1.
CC PROSITE; PS00059; ADH_ZINC; 1.
CC Oxidoreductase; Zinc; Metal-binding; NADP; Lignin biosynthesis.
FT METAL 47 47 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 69 69 ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL 100 100 ZINC 2 (BY SIMILARITY).
FT METAL 103 103 ZINC 2 (BY SIMILARITY).
FT METAL 106 106 ZINC 2 (BY SIMILARITY).
FT METAL 114 114 ZINC 2 (BY SIMILARITY).
FT METAL 163 163 ZINC 1 (CATALYTIC) (BY SIMILARITY).

```

SQ SEQUENCE 360 AA; 39129 MW; 28F1D980B08C4096 CRC64;

Query Match 39.3%; Score 691.5; DB 1; Length 360;
Best Local Similarity 44.0%; Pred. No. 2.1e-49;
Matches 150; Conservative 52; Mismatches 100; Indels 39; Gaps 9;

```

QY 1 ETGRTDVRFKVLYGVCHSDIHMAKNKMGTSYPIVPGHGLGVVTVGCKYKKEKSWRQ 60
Db 31 ETGPEDEVFIKIYICGICHTDIHQTKNDLGASNYPMVPGHEVVGVGVSDVTKPK- --V 87
QY 61 GRCWLHGRURPTCENCIIHLENYCPNLIOTYGSKYVDGTMVYGGYSNNMVVDHFTV 117
Db 88 GDCVGDGTIVGCKTCRCADVEQYCNKKIWSYNDVYTDGKPTGGFSGHVVVDQKFW 147
QY 118 RIPNPLDGAAPLLCAGITTYSPWRYYGLDK-FGMHLGVWPRFRSRPLNLPGLMGS 176
Db 148 KIPDMAPEQAAPLLCAGITTYSPWRYYGLDK-FGMHLGVWPRFRSRPLNLPGLMGS 196
QY 177 RLQSLVPLPLKEGSGYGT-----SPALVH-----SLARTDQDOMEAAMSTMDGLID 222
Db 197 GVK-----LAKAMGHVTVISSDKKKEEAIDHLGADAYLVSSDATQCEAADSLOYIID 251
QY 223 TVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLVPFPLI-IGRMVAGSAIGMKETQM 281
Db 252 TVPVFHPLEPVLSLKLGKLLILMGVINTPLQF-ISPVMVLGRKAITGSGFSGMKEETEM 310
QY 282 IDFAAENHITADIEVIPIDYLTAMERVVKXDVFRFVIDV 322
Db 311 LDFCNKGITSTIEVWQMDYINTAFERLEKNDVRYRFVVDV 351

```

Search completed: July 6, 2004, 13:37:17

Job time : 9.38046 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2004, 13:33:14 ; Search time 31.8458 Seconds
(without alignments)
3229.910 Million cell updates/sec

Title: US-09-857-518A-31
Perfect score: 1758
Sequence: 1 ETGATDVRFKLYGCVCHSD.....ERVVKDVRFRVIDVENTL 326

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1064	60.5	359	10 Q9ATW1	Q9atw1 fragaria an
2	984	56.0	362	10 Q94G59	Q94g59 populus tre
3	970	55.2	363	10 Q7XAB2	Q7xab2 camptotheca
4	950.5	54.1	359	10 Q8H018	Q8h018 solanum tub
5	936	53.2	362	10 Q8L7U8	Q8l7u8 populus tre
6	881	50.1	360	10 Q94K02	Q94k02 arabidopsis
7	876	49.8	360	10 Q8L884	Q8l884 arabidopsis
8	853	48.5	363	10 Q65621	Q65621 arabidopsis
9	839	47.7	370	10 Q8S411	Q8s411 lolium pere
10	819.5	46.6	375	10 Q8S411	Q8s411 arabidopsis
11	802	45.6	376	10 Q9SU25	Q9suj25 arabidopsis
12	790	44.9	420	10 Q7X898	Q7xe898 oryza sativ
13	784.5	44.6	289	10 Q9M722	Q9m722 lycopersico
14	755.5	43.0	360	10 Q7XW00	Q7xw00 oryza sativ
15	712.5	40.5	391	10 Q8H809	Q8h809 oryza sativ
16	707	40.2	410	10 Q7XLD5	Q7xld5 oryza sativ

17	679	38.6	357	10 Q9M632	Q9m632 populus tre
18	675	38.4	357	10 Q9FSC7	Q9fsc7 populus tri
19	666	37.9	407	10 Q8S412	Q8s412 lolium pere
20	662	37.7	361	10 Q947S1	Q947s1 festuca aru
21	660	37.5	361	10 Q947S3	Q947s3 festuca aru
22	660	37.5	361	10 Q947S2	Q947s2 festuca aru
23	654	37.2	361	10 Q947S0	Q947s0 festuca aru
24	648	36.9	356	10 Q9FUN8	Q9fun8 eucalyptus
25	641	36.5	335	10 Q8W420	Q8w420 medicago sa
26	626.5	35.6	348	2 Q9AE96	Q9ae96 mycobacteri
27	622.5	35.4	350	16 Q88K65	Q88k65 pseudomonas
28	622.5	35.4	350	16 Q884B3	Q884b3 pseudomonas
29	622	35.4	349	2 Q8KRC3	Q8krc3 myxococcus
30	621.5	35.4	352	5 Q9ULF0	Q9ulf0 leishmania
31	621	35.3	347	16 Q92MD4	Q92md4 rhizobium m
32	616.5	35.1	355	16 Q8XQ04	Q8xq04 raietonia s
33	614	34.9	365	10 Q8L9U1	Q8l9ul arabidopsis
34	613.5	34.9	347	16 Q82144	Q82144 streptomyce
35	613.5	34.9	362	16 Q9CBQ3	Q9cbq3 mycobacteri
36	612	34.8	321	10 Q04079	Q04079 zinnia eleg
37	610.5	34.7	354	10 Q8H859	Q8h859 oryza sativ
38	609	34.6	368	16 Q8UF43	Q8uf43 agrobacteri
39	596.5	33.9	349	16 Q06007	Q06007 bacillus su
40	596	33.9	358	16 Q882D1	Q882d1 pseudomonas
41	588.5	33.5	346	16 Q8ES57	Q8es57 oceanobacil
42	587.5	33.4	355	10 Q9CAI3	Q9cai3 arabidopsis
43	583.5	33.2	352	16 Q8PRD2	Q8prd2 xanthomonas
44	581.5	33.1	353	16 Q911J9	Q911j9 pseudomonas
45	580.5	33.0	352	16 Q8Y1Q7	Q8y1q7 ralstonia s

ALIGNMENTS

RESULT 1

Q9ATW1 PRELIMINARY; PRT; 359 AA.
ID Q9ATW1
AC Q9ATW1;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Cinnamyl alcohol dehydrogenase.
OS Fragaria ananassa (Strawberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids I; Rosales; Rosaceae; Rosoideae; Fragaria.
OX NCBI_TaxID=3747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Chandler;
RA Blanco-Portales R.R., Caballero-Capullo J.J., Munoz-Blanco J.J.;
RT "Cloning, expression and immunolocalization pattern of a cinnamyl
alcohol dehydrogenase gene from strawberry (Fragaria x ananassa c.v.
Chandler).";
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
DR EMBL; AF320110; RAX38509.1;
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002328; Adh_zinc.
DR InterPro; IPR02085; Adh_zn family.
DR InterPro; IPR00345; CytC_heme_BS.
DR Pfam; PF00107; Adh_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
DR PROSITE; PS00130; CYTOCHROME_C; 1.
KW Metal-binding; Oxidoreductase; Zinc.
SQ SEQUENCE 359 AA; 38907 MW; 5452F249A8ECAA66 CRC64;

Query Match 60.5%; Score 1064; DB 10; Length 359;
Best Local Similarity 64.5%; Pred. No. 6.1e-89;


```
AC Q8H0L8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alcohol NADP+ oxidoreductase.
GN DRD-1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bintje;
RA Montesano M., Hyttinen H., Wettstein R., Palva T.E.;
RT "A novel potato defense-related alcohol:NADP+ oxidoreductase induced
in response to Erwinia carotovora."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ439933; CAD29291.1; -
DR GO: GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000345; CytC_heme_BS.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 359 AA; 3938 MW; AA6B32922E2B73A66 CRC64;

Query Match 54.1%; Score 950.5; DB 10; Length 359;
Best Local Similarity 54.2%; Pred. No. 1.5e-78;
Matches 186; Conservative 56; Mismatches 62; Indels 39; Gaps 6;

QY 2 TGAATDVRKLYCGVCHSDIHMANKNDWGTSTYPIVPGHELGVVTVGVCKVKFK-SWRQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 TGEKHQVKVYCGVCHSDIHLQKNEGNTKPMVPGHGVVGVVGVKVEKVGDKV 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 GRCWLHGRLEPTCENCIHLENCNLIQTYGKYDGTMTYGGYSNNMVTDEHFIVRIP 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 95 GVGCMVGSCE-KCENCTVDLENYCPRIPTYNGYSLDGLTFGYSNNMVSDEHFVVRP 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPMHGVWPRFRSRPPLNLPGLWGSRLQS 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 ENLSMD-AALLCAGITTYSPKTYFLGLDKPMHGV-----VGLGG----- 193
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 LVPLIKEGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDT 223
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 194 LGHMAVKFAKAFGKTVISTSAKKQEAERLGAOSFLISRDPEQKAAAMTLDGIIDT 253
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 224 VPAVRLEPLISLKTNGKVVTVGIAVQPLDLPVFPFLLIGRKVAGSAIGMKETQEMID 283
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 254 VSAVHPILPLLMKSHGKLVWVGAPEKPVLPVFPFLLMGRKLVAGSCIGMKETQEMID 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 284 SAEHNITADIEVIPIDYLTNTAMERVVKDVRFVFDVENTL 326
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 314 FRAKHNITADIEVPMIDYVNTLRLLKSDVKRFVLDIGNTL 356
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
Q8L7U8
ID Q8L7U8 PRELIMINARY; PRT; 362 AA.
AC Q8L7U8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative sinapyl alcohol dehydrogenase.
OS Populus tremula x Populus tremuloides.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosidia I; Malpighiales; Salicaceae; Salix; Populus.
OX NCBI_TaxID=47664;
```

```
RP SEQUENCE FROM N.A.
RA Israelsson M., Eriksson M.E., Hertzberg M., Aspeberg H., Nilsson P.,
RA Moritz T.;
RT "Changes in gene transcription in the wood-forming tissue of
transgenic hybrid aspen with increased secondary growth."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY126444; ANW95578.1; -
DR GO: GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000345; CytC_heme_BS.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 1.
SQ SEQUENCE 362 AA; 39527 MW; BA512D2CEFC97EC6 CRC64;

Query Match 53.2%; Score 936; DB 10; Length 362;
Best Local Similarity 52.2%; Pred. No. 3.3e-77;
Matches 179; Conservative 54; Mismatches 74; Indels 36; Gaps 3;

QY 1 ETGATDVRKLYCGVCHSDIHMANKNDWGTSTYPIVPGHELGVVTVGVCKVKFKSWRQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 34 DNGVEDVTIKLYCGVCHSDIHAANKNGWFSRYPLVPGHEIVGIVTKIGSNVKKFKVDDQ 93
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 GRCWLHGRLEPTCENCIHLENCNLIQTYGKYDGTMTYGGYSNNMVTDEHFIVRIP 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 94 VGVGMVNSCKSCBYCDQSENFCFMFTYNAQNYDGTXYGYSNTIVVDQHFVLRIP 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 DNLPLDGAAPLLCAGITTYSPWRYGLDKPMHGVWPRFRSRPPLNLPGLWGSRLQS 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 154 DSMFADGAPLLCAGITTYSPKTYFLGLDKPMHGV-----VGLGG----- 194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 LVPLIKEGSYGTSPALMHS-----LLRTDQDQMEAAAMSTMDGIIDT 223
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 195 LGHMAVKFAKAFGKTVISTSAKKQEAERLGAOSFLISRDPEQKAAAMTLDGIIDT 254
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 224 VPAVRLEPLISLKTNGKVVTVGIAVQPLDLPVFPFLLIGRKVAGSAIGMKETQEMID 283
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 VSAVHALAPLLSLKTNGKLVTLGLPEKLELPFLVPLVGRKLVGSGDGGMKETQEMLD 314
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 284 SAEHNITADIEVIPIDYLTNTAMERVVKDVRFVFDVENTL 326
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 325 FCAKHNITADIEVIRMDQINTAMDRLAKSDVRFVFDVANSL 357
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
Q94K02
ID Q94K02 PRELIMINARY; PRT; 360 AA.
AC Q94K02;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cinnamyl-alcohol dehydrogenase CAD1 (Putative alcohol dehydrogenase)
DE (EC 1.1.1.195).
GN T22F8.230 OR CAD9 OR AT4G39330.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosidia II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Bach J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinozaki K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
```

	CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
	OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.
	OX	NCBI TaxID=3702;
	RN	[1]
	RP	SEQUENCE FROM N.A.
	RA	Haas B.J., Volkovsky N., Town C.D., Troupkhan M., Alexandrov N.,
	RA	Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
	RT	"Full-length messenger RNA sequences greatly improve genome
	RT	annotation.";
	RL	Genome Biol. 0:0-0(2002).
	RZ	[2]
	RP	SEQUENCE FROM N.A.
	RA	Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
	RA	Feldmann K.;
	RT	"Full-length cDNA from Arabidopsis thaliana.";
	RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
	DR	EMBL; AY087363; AA064913.1; -
	DR	GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
	DR	GO; GO:0005489; F:electron transport activity, zinc-
	DR	GO; GO:0008270; F:zinc ion binding; IEA.
	DR	GO; GO:0006118; P:electron transport; IEA.
	DR	InterPro; IPR002328; ADH_zinc.
	DR	InterPro; IPR002085; Adh_zn family.
	DR	InterPro; IPR000345; Cytochrome BS.
	DR	Pfam; PF00107; ADH_zinc_N; 1.
	DR	PROSITE; PS00059; ADH_ZINC; 1.
	DR	PROSITE; PS00190; CYTOCHROME_C_1.
	SQ	SEQUENCE 360 AA; 38933 MW; 06ED3B0581785759 CRC64;
		Query Match 49.8%; Score 876; DB 10; Length 360;
		Best Local Similarity 48.9%; Pred. No. 1e-71;
		Matches 171; Conservative 52; Mismatches 77; Indels 50; Gaps 5;
Qy	1	ETGADTVRFKVLVCGVCHSDTHMAKNWDGTSTYTPVGHELVGVVTEGCKVKPKSMRQ 60
Dd	34	DNGENDVTKLFGCVCHTDLHTINDGWYSYFVPVGHEIVGATKVGNWTKFKEGDR 93
Qy	61	-----GRCLHLRLPTCENCIHHLENYCPNIQTGYSKYDGTMTYGGSYNNMVTDE 113
Dd	94	VGVGVISGSC-----QSCESCDDLENYCQMSFTNAIGSDGTKNYGGSYSENVVDQ 146
Qy	114	HFIVRIPDNLPDCAAPLLCAGITTYSPRWYGLDKPMHIGVEMPRFRSRPPNLPLGL 173
Dd	147	RFVLRFEPNLFPSDSRAPLLCAGIIVTSFMKYKYGWTEAKHLGV-----AGL 192
Qy	174	WGSXLQSLVPELIIEGSGSYGTFSPALMHSS-----LIERTDQMEAAAVST 216
Dd	193	GG-----LGHWVWGKGAFLKVVTVISSSTSKABEAINHLGADGFLVTTDPQMKAAIGT 247
Qy	217	MDGIIDTVPAVRPERPLISLLKTNGKKVTVGIAVQPLDLFPPELLIGRKWAGSAGGMK 276
Dd	248	MDYIIDTISAVHALYPILLGLLKNGKLIAGLPEKPELPNFPIVLGRKNYGGSDVGMMK 307
Qy	277	ETQMIDFAAEHNITADIEVIPDYLTAMERVVKQVRFREVIDENTL 326
Dd	308	ETQMLDFCAKHNTADIEIKMDEINTAMERLAKSIVRYEFVINVAANSL 357

RESULT 8	
065621	
ID 065621	PRELIMINARY.
	DDT- 363 AA

DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Cinnamyl alcohol dehydrogenase-like protein, subunit A (Cinnamyl
DE alcohol dehydrogenase-like protein, LCADA) [putative alcohol
DE dehydrogenase] (EC 1.1.1.195).
GN LCADA OR F20D10.90 OR AT4G37970 OR CAD6.
OC Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi


```

QY 113 EHFIVRIPDNLPGDGAAPLLCAGITTYSPWYYGLDK-PGMHLGVEMPRFRSRPPLNLP 171
Db 141 HRFVLSIPDLPSDGAAPLLCAGITTYSPWYYGLDK-PGMHLGVEMPRFRSRPPLNLP 186
QY 172 GLWGRSLQSLVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAM 214
Db 187 GLGG-----LGHIAVAKGAFGLRVTVISRSSEKEREADLGDGDFLVTDSQKMKEAV 241
QY 215 STMGGIITVPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGG 274
Db 242 GTMDFIIDVSAHALLPLFSLKVNKGLVALGLPEKPLDPLFSLVGLGRKMWGSGQIGG 301
QY 275 MKETQEMIDFAAEHNITADIEVIPIDYLNMTAMERVVKDVRFRFVIDVENTIL 326
Db 302 MKETQEMLEFCAKHIVSDIELIKMSDINSMDRLAKSDVRFRFVIDVANSI 353

RESULT 12
QYX98
ID QYX98 PRELIMINARY; PRT; 420 AA.
AC QYX98;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Putative cinnamyl alcohol dehydrogenase.
GN OSUNBA0073D04.1.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017097; AAP53892.1; -
SQ SEQUENCE 420 AA; 43979 MW; 11204EE1143288A9 CRC64;

Query Match 44.9%; Score 790; DB 10; Length 420;
Best Local Similarity 46.9%; Pred. No. 9.6e-64;
Matches 161; Conservative 56; Mismatches 88; Indels 38; Gaps 5;

QY 2 TGAIDVRFKVLYCGVCHSDIIMKNDMCTSTYPLVPGHGLGVNTVGCCKVKKSW-RQ 60
Db 92 TGDDVAIKILFCGICHSDLCIKNEWKHSIYPLVPGHGLGVNTVGCCKVKKSW-RQ 151
QY 61 GRCLWHLRPTCSNCHLHLENYCNLTQYSGKYDGTMTYGGYNNMTDEHFIVRIP 120
Db 152 GVGCMVNSCR-SCSSCNNGFENHCEPGVFTVNSVDKGTVTYGGYSSVWVHERFVWFP 210
QY 121 DNLPLDGAAPLLCAGITTYSPWYYGLDK-PGMHLGVEMPRFRSRPPLNLPGLWGRSLQ 180
Db 211 EAMPLDVGAPLLCAGITTYTPMKYHGLNAPKRGV-----LGLGGLG----- 253
QY 181 LVPLIKEGSGYGTSPALMHS-----LLRTDQDQMEAAAMSTWDGIIDT 223
Db 254 --HVAVFAFAFGKLVTVISSPPCKKREALERLGADAFVWSSSABEMAEASTWDGIVNT 311
QY 224 VPAVRPLEPLISLKTNGKVVTVGIAVQPLDLPVPLIIGRKWVAGSAIGGKETQEMID 283
Db 312 VSANTPMAPVIALKPKNGKMLVGLPENPLEVFPFSLVGHGRTLAGSNGIGMAQTQEMIE 371
QY 284 FAEHNITADIEVIPIDYLNMTAMERVVKDVRFRFVIDVENTIL 326

```

```

Db 372 LAAGHGVTAIDIEVIGADDVNTAMERLAKADVRFRFVIDVGNL 414

RESULT 13
QYX722
ID QYX722 PRELIMINARY; PRT; 289 AA.
AC QYX722;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BL13 (Fragment).
GN BL13.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Rio Grande;
RA Thimmony R.L.; Martin G.B.;
RT "Rapid induction of a Novel Gene and an eli3 Homolog During the Pto-
mediated Resistance Response in Tomato."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF146691; AAF72100.1; -
DR GO; GO:0004024; F.alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0008270; F.zinc ion binding; IEA.
DR InterPro; IPR002085; Adh zn family.
DR Pfam; PF00107; Adh_zinc_N; 1.
FT NON TER 1
SQ SEQUENCE 289 AA; 31333 MW; B2BAFDC316A66CE8 CRC64;

Query Match 44.6%; Score 784.5; DB 10; Length 289;
Best Local Similarity 52.4%; Pred. No. 1.8e-63;
Matches 161; Conservative 50; Mismatches 57; Indels 39; Gaps 6;

QY 33 GHELVGVVTEGCKVKKPK-SWRQRCWMLHGLRPTCEHLENYCNLTQYSGKY 96
Db 1 GTRVVGWVTEGSKVKKPKGVKGVGCVGWSGR-KCNCSDVLENYCPRQIPTNGYSL 59
QY 97 DGTMTYGGYNNMTDEHFIVRIPDNLPGDGAAPLLCAGITTYSPWYYGLDKPGMHLGV 156
Db 60 DGTMTYGGYSDVWVSDEHFVVRWPELNSD-AAPLLCAGITTYSPKXPLDKPGMHLGV 118
QY 157 EWRFRFRSRPPLNLPGLWGRSLQSLVPLIKEGSGYGTSPALMHS----- 231
Db 119 -----VGLGG-----LGHMAVFAFAFGTKVTVISTANKKKEAERLGAJ 159
QY 202 --LLRTDQDQMEAAAMSTWDGIIDTVPVPRLEPLISLKTNGKVVTVGIAVQPLDLPVP 259
Db 160 SPLISRDPEQKAAAMNTLDGIIDTVSVVHPLPLMLMKSHGKLWVWGAPEKEVELPVP 219
QY 260 LIIGRKWVAGSAIGGKETQEMIDFAAEHNITADIEVIPIDYLNMTAMERVVKDVRFRFV 319
Db 220 LLMGKRLVAGSCIGGKETQEMLEFAAEHNITPDIEVPMVEYVNTALERLLKSDVKYRFV 279
QY 320 IDVENTIL 326
Db 280 LDIGNL 286

RESULT 14
QYXWUO
ID QYXWUO PRELIMINARY; PRT; 360 AA.
AC QYXWUO;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE OSUNBA0065B15.11 protein.
GN OSUNBA0065B15.11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```




Creation date: 07-14-2004
Indexing Officer: MMOHAMMED1 - MENEN MOHAMMED
Team: OIPEScanning
Dossier: 09857518

Legal Date: 07-09-2004

No.	Docode	Number of pages
1	SRNT	26

Total number of pages: 26

Remarks:

Order of re-scan issued on